



**This electronic thesis or dissertation has been
downloaded from Explore Bristol Research,
<http://research-information.bristol.ac.uk>**

Author:

Mora Rodriguez, Maria

Title:

**Research into the possible effects of new and emerging information technology on the
delivery and consumption of environmental information in businesses**

General rights

Access to the thesis is subject to the Creative Commons Attribution - NonCommercial-No Derivatives 4.0 International Public License. A copy of this may be found at <https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>. This license sets out your rights and the restrictions that apply to your access to the thesis so it is important you read this before proceeding.

Take down policy

Some pages of this thesis may have been removed for copyright restrictions prior to having it been deposited in Explore Bristol Research. However, if you have discovered material within the thesis that you consider to be unlawful e.g. breaches of copyright (either yours or that of a third party) or any other law, including but not limited to those relating to patent, trademark, confidentiality, data protection, obscenity, defamation, libel, then please contact collections-metadata@bristol.ac.uk and include the following information in your message:

- Your contact details
- Bibliographic details for the item, including a URL
- An outline nature of the complaint

Your claim will be investigated and, where appropriate, the item in question will be removed from public view as soon as possible.

**Research into the possible effects of new and
emerging information technology on the delivery and
consumption of environmental information in
businesses**

Maria Mora-Rodriguez

Dissertation submitted to the University of Bristol in accordance with the requirements for award
of the degree of Engineering Doctorate in the Faculty of Engineering

Systems Centre
Queens School of Engineering
University of Bristol
August 2018

Abstract

One of the greatest challenges faced by modern businesses is being able to achieve industrial progress while reducing environmental impact. Paradoxically, it is also one of the largest opportunities for businesses in a global economy.

There is a growing belief among investors, governments, and the public at large that organisations have the responsibility to account for their impact on the environment and society. This has resulted in companies publishing documents such as corporate social responsibility or sustainability reports that disclose financial, social, environmental and governance information to their stakeholders. However, the quality of the non-financial information contained in such reports is varied and often considered of questionable value.

This thesis focuses on (1) what can be done to encourage more robust corporate reporting to drive positive environmental change in companies, and (2) how better-informed decisions can be facilitated in future investments, policy-making and consumption.

This thesis is framed in an EngD programme and presents three types of studies:

- (1) Corporate reporting and the effectiveness of disclosure based on prior work study research and qualitative market research with stakeholders;
- (2) Envisioning how to evolve corporate reporting so that environmental issues are considered in decision making, thereby allowing current challenges to be overcome and creating new opportunities;
- (3) Solution-driven studies for building an efficient and useful corporate reporting template using Semantic Web technologies, reporting standards and Artificial Intelligence techniques.

This research contributes to knowledge in several ways: it presents insights from the market; identifies key corporate reporting features (drivers, constraints, improvements and impacts); proposes best practices; and builds technical solutions. Much of the work focuses on tackling the challenges of better data standardisation, data connectivity, and data integration with applications in complex corporate information. The intention is to provide material to promote the implementation of this work and further research.

The results of this thesis are also providing the groundwork for a number of corporate reporting working groups and task forces led by organisations like the Institute of Chartered Accountants in England and Wales (ICAEW), Financial Reporting Council (FRC), Carbon Disclosure Project (CDP), The Spanish Association of Accountants (AECA), XBRL International, XBRL Europe and XBRL Spain. At the same time, we are encouraging a closer collaboration between reporting initiatives and technical communities, as well as the evolution of reporting standards, such as XBRL standards, to better support environmental reporting needs.

Funding

This work was supported by the Systems Centre at the University of Bristol, the EPSRC funded Industrial Doctorate Centre in Systems (Grant EP/G037353/1) and CDP Worldwide, London, UK.

Research for the paper presented in Chapter seven, ‘The role of XBRL in overcoming climate-related reporting challenges’, was funded by the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 649982.

Author's Declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

Signed.....Maria Mora-Rodriguez

Date 25/08/2018

A handwritten signature in blue ink, appearing to read 'Maria', with a stylized flourish underneath.

Table of Contents

Chapter 1. Introduction	1
1.1 Background and motivation	1
1.2 About the industrial sponsor of this EngD.....	4
1.3 Purpose of this research project	6
1.4 Thesis structure.....	7
Chapter 2. Methodologies	12
2.1 Qualitative research methodology.....	12
2.2 Taxonomy design methodology	15
2.3 Machine Learning methodology	16
Chapter 3. Context	18
3.1 Background	18
3.2 Voluntary and mandatory reporting schemas	21
3.3 Corporate reporting and market behaviour	22
3.4 Use of information systems	26
3.5 Corporate reporting initiatives.....	27
3.6 Semantic web and linked open data in the corporate world	30
3.7 Research goals	34
Chapter 4. Literature review and related work.....	36
4.1 About CDP and its theory of change	37
4.2 CDP in the literature	38
4.2.1 CDP strategy.....	38
4.2.2 The influence of CDP through its reporting initiative	38
4.2.3 The influence of CDP data approach to create impact in the market	39
4.3 The role of information systems in environmental sustainability	40
4.4 Corporate reporting over the internet	42
4.5 Technologies for corporate reporting: XBRL and Linked Data.....	44
4.5.1 Data Standardisation - XBRL.....	44
4.5.2 Data connectivity - Linked Data	47
4.5.3 Standardisation and Data Connectivity working together	49
Chapter 5. Stakeholders' study: the needs for achieving environmental sustainability	52
5.1 Introduction	52
5.2 Background	54
5.2.1 CDP as an environmental reporting system	55

5.3 Research questions	56
5.4 Results.....	60
5.4.1 Drivers.....	60
5.4.2 Constraints	68
5.4.3 Improvements.....	71
5.4.4 Impacts.....	73
5.5 Discussion	74
5.5.1 Contribution to Elliot's model	74
5.5.2 Contrasts with previous findings on CDP literature	76
5.5.3 What can be learned from CDP	77
5.6 Conclusions	79
Chapter 6. The role of interoperability in sustainability decisions	82
6.1 Introduction	83
6.2 Background	84
6.2.1 Standardisation and interoperability	84
6.2.2 The case of CDP, an environmental reporting initiative and a source of environmental data	85
6.3 Research questions	86
6.4 Results.....	87
6.4.1 Legal interoperability	89
6.4.2 Organisational interoperability	90
6.4.3 Semantic interoperability.....	91
6.4.4 Technical interoperability	93
6.5 Discussion	95
6.6 Conclusions	98
Chapter 7. Influencing the practices of XBRL within CDP.....	100
7.1 The role of XBRL in overcoming environmental-related reporting challenges	104
7.1.1 Introduction	104
7.1.2 Task force on climate-related financial disclosures (TCFD)	105
7.1.3 The role of technology: overcoming environmental-related reporting challenges through XBRL.....	106
7.1.4 Conclusions	113
7.2 Using XBRL for representing the CDP environmental data: insights and modelling decisions .	114
7.2.1 Motivation	114

7.2.2 Data modelling: 2016 climate change questionnaire technical requirements	114
7.2.3 Modelling CDP concepts in XBRL.....	122
7.2.4. Style guide and naming convention	141
7.2.5 Taxonomy validation	151
7.3 Conclusions	152
Chapter 8. Adopting Semantic Technologies for Effective Corporate Transparency	154
8.1 Introduction	155
8.2 XBRL fundamentals relevant for this chapter.....	158
8.3 Transforming XBRL into linked data	161
8.3.1 Lightweight vocabulary for XBRL (XBRL)	161
8.3.2 From XBRL data to Linked data	164
8.3.3 Linking XBRL data to other data	168
8.4 Validation.....	169
Goal 1. Data coverage using DBpedia	170
Goal 2. Cross data source analysis using CNMV and CDP data.....	171
Goal 3. Data accuracy using DBPedia and CNMV data	171
8.5 Discussion	171
8.6 Conclusions and further research	173
Chapter 9. A Solution To Aligning Corporate Reporting Frameworks: The Case of GRI and CDP... 175	
9.1 Introduction	175
9.2 Problem situation.....	178
9.4 Method step by step	180
9.5 Results.....	184
9.6 Discussion and conclusions	186
Chapter 10. Conclusions and further work	189
10.1 The problem situation and vision.....	191
10.2 Solutions	196
10.3 Impact and contributions of this research	197
10.4 Further work	198
11.References	Error! Bookmark not defined.
Annex A - Mandatory and voluntary disclosure initiatives	217

List of Figures

Figure 1. Corporate reporting ecosystem	3
Figure 2. Problems and challenges in corporate reporting processes.....	4
Figure 3. System map: Exploring CDP as a network of actors	5
Figure 4. Thesis structure.....	11
Figure 5. Stakeholders and market map	13
Figure 6. Taxonomy Design methodology.....	15
Figure 7. Text classification methodology	17
Figure 8. The STOXX Sustainability index against a general stock market index	19
Figure 9. Openness score at the UK open data portal data.gov.uk [Accessed August 2019]	33
Figure 10. Belief-Action-Outcome (BAO) Framework for IS Research on Sustainability	41
Figure 11. Conceptual model of IS and ecological sustainability.....	42
Figure 12. Corporate dialogue as an incremental step from traditional online disclosure	43
Figure 13. Linking Open Data cloud diagram 2017, by Andrejs Abele, John P. McCrae, Paul Buitelaar, Anja Jentsch and Richard Cyganiak	48
Figure 14. A socio-technical system to impact on environmental sustainability by Elliot [97].....	54
Figure 15. Level of reporting demands per market.....	79
Figure 16. Dissemination of XBRL work within relevant industrial forums. (European Central Bank, 2017).....	103
Figure 17. XBRL implementation within the European System of Financial Supervision	107
Figure 18. Benefits of XBRL	109
Figure 19. 2016 Climate Change entry points	115
Figure 20. Climate Change questionnaire structure.....	116
Figure 21. Question pathway	120
Figure 22. Example of enumeration structure in CDP questionnaire	121
Figure 23. Example of attached document structure.....	122
Figure 24. Climate Change Taxonomy physical architecture.....	135
Figure 25. Enumeration concept - list of values definition.....	138
Figure 26. Formulas - example of existence rule definition	139
Figure 27. Formulas- example of value assertion definition	140
Figure 28. Fact representation in XBRL	159
Figure 29. Classes and relationships XBRL Ontology	164
Figure 30. RDF representation of a simple XBRL fact	165
Figure 31. RDF representation of a tuple fact	166
Figure 32. RDF representation of a dimensional fact	167
Figure 33. DGI representation in RDF using FOAF vocabulary	167
Figure 34. From XBRL to Linked data architecture. Inputs, outputs and technologies.....	170
Figure 35. CDP responses classified by question number	181
Figure 36. CDP data in 2-dimensional space	182
Figure 37. Total pages of GRI reports classified per CDP question number and confidence levels... 185	
Figure 38. Number of GRI pages and associated confidence level.....	186
Figure 39. Levels of assurance in auditing. Source: https://goo.gl/Ka94Tz [Accessed August 2019] 194	

Figure 40. Annual report and the different level of assurance	195
Figure 41. Technical problems in environmental reporting	197
Figure 42. Disciplines that can solve corporate reporting problems.....	199

List of Tables

Table 1. Distribution of participants per stakeholder group	14
Table 2. List of interview questions	59
Table 3. Classification of the results per interview coding	64
Table 4. Interview results and interoperability properties.....	88
Table 5. External sources to link with CDP data	93
Table 6. Climate Change Questionnaire 2016 (common module skeleton)	118
Table 7. Oil and Gas sector specific module skeleton	119
Table 8. Correspondence between presentation linkbase and Core module framework	126
Table 9. Correspondences between presentation linkbase and Oil and Gas module framework.....	126
Table 10. Definition linkbase – Common module	129
Table 11. Extensible enumerations within the definition linkbase	132
Table 12. Definition linkbase - Oil and Gas module	133
Table 13. Reference linkbase	134
Table 14. Data types and units.....	141
Table 15. Data coverage: Information about the context of Repsol S.A.	171
Table 16. Data analysis: Emission intensity of Repsol S.A. in 2015	171
Table 17. Data consistency: Reliability of Equity figure presented in DBpedia	171
Table 18. Convergence in the area of Climate Change risks between CDP and GRI frameworks.....	177
Table 19. SVM results: precision, recall and score	183
Table 20. Results: GRI reports classified in CDP questions.....	184

Glossary

AECA	Asociación Española de Contabilidad y Administración de Empresas / The Spanish Association of Accountants
AICPA	American Institute of Certified Public Accountants
API	Application programming interface
BPB	Best Practices Board
CEN	European Committee for Standardization
CDP	Carbon Disclosure Project
CNMV	Comisión Nacional del Mercado de Valores / Spanish Security Exchange Commission
DJSI	Dow Jones Sustainability Index
EBRC	Enhanced Business Reporting Consortium
EDINET	Electronic Disclosure for Investors Network
EMAS	EU Eco-Management and Audit Scheme
ESG	Environmental, Social and Governance
GAAP	Generally Accepted Accounting Principle
GRI	Global Reporting Initiative
GHG	Greenhouse gas
IFRS	International Financial Reporting Standards
IIRC	International Integrated Reporting Council
ILO	International Labour Organization
ISAR	International Standards of Accounting and Reporting
LOD	Linked Open Data
LOV	Linked Open Vocabularies
NAZCA	Non-state Actor Zone for Climate Action
OIM	Open Information Model
RDF	Resource Description Framework
SASB	Sustainability Accounting Standards
SNSs	Social Network Sites
SPARQL	Protocol and RDF Query Language
URIs	Uniform Resource Identifier
XBRL	eXtensible Business Reporting Language
XML	eXtensible Markup Language
XSB	XBRL Standards Boards
W3C	World Wide Web Consortium

List of academic publications and conferences

- 1) Mora-Rodriguez, M., Auguste-Atemezing, G. and Preist, C. (2017). Adopting Semantic Technologies for Effective Corporate Transparency. *14th European Semantic Web Conference*. (pp 655-670). *Published by Springer International Publishing*. DOI: 10.1007/978-3-319-58068-5_40

Paper written by Maria Mora, with revisions from Chris Preist and Ghislain Auguste (Modeca Lab Research). Presented by Maria Mora at *the 14th European Semantic Web Conference* (Portoroz, Slovenia). June 2017. Presentation available here: <https://www.slideshare.net/MariaMora69/adopting-semantic-technology-for-effective-corporate-transparency>

- 2) Mora-Rodriguez, M. and Faria, P. (2017). The role of XBRL in overcoming climate-related reporting challenges. *Published at the CDP website*. (<http://bit.ly/2lBOMop>)

Paper was written by Maria Mora, with revisions from Pedro Faria (CDP), John Turner (XBRL International), Wada-San (XBRL Asia), Derek de Brandt (Eurofiling), Jose Luis Lizcano (AECA) and Ignacio Boixo (European Committee for Standardisation-CEN). Presented by Maria Mora at the 18th XBRL Europe Day (European Central Bank, Frankfurt, Germany). June 2016.

This work was funded by European Union's Horizon2020 research and innovation programme under grant agreement No 649982.

- 3) Rezaee, Z., Homayoun, S. and Mora-Rodriguez, M. (2017) "Integration of real-time analysis of big data into sustainability attributes". In CEUR Workshop Proceedings (Vol. 1890) - XBRL Academic Track Proceedings. <http://ceur-ws.org/Vol-1890/paper13fullPaper.pdf>

- 4) Mora-Rodriguez, M. and Preist, C. (2016). The role of interoperability in sustainability decisions: the case of CDP. *4th International Conference on ICT for Sustainability*. (pp 219-228) *Published by Atlantis Press*. [doi:10.2991/ict4s-16.2016.26](https://doi.org/10.2991/ict4s-16.2016.26)

Paper was written by Maria Mora, with revisions from Chris Preist. Presented by Maria Mora at *the 4th International Conference on ICT for Sustainability* (Amsterdam, Netherland). August 2016.

- 5) Mora-Rodriguez, M. and Hicks, I. (2016). A solution to align corporate reporting frameworks: The case of GRI and CDP. *1st Data amplified-XBRL International Conference*.

Paper was written by Maria Mora, with revisions from Ian Hicks (XBRL International). Presented by Maria Mora at *the 1st Data Amplified Conference* (Singapore, Malaysia). November 2016. Presentation available here: <https://www.slideshare.net/MariaMora69/a-solution-to-align-corporate-reporting-frameworks-the-case-of-gri-and-cdp>

- 6) Mora-Rodriguez, M. and Preist, C. (2015). Towards the effectiveness of global markets: the convergence between US GAAP and IFRS through XBRL and linked data. *19th Annual Financial Reporting and Business Communication Conference*.

Paper was written by Maria Mora, with revisions from Chris Preist. Presented by Maria Mora at *the 19th Annual Financial Reporting and Business Communication Conference*. (University of Bristol, United Kingdom). July 2015.

- 7) Mora-Rodriguez, M. and Preist, C. (2014). The practice of system dynamics: exploring the role of XBRL in an environmental reporting system. *33rd World Continuous Auditing & Reporting Symposia*.

Paper was written by Maria Mora, with revisions from Chris Preist. Presented by Maria Mora at *33rd World Continuous Auditing & Reporting Symposia*. (Rutgers University – Newark, USA). November 2014. Presentation available here: <https://www.slideshare.net/MariaMora69/the-practice-of-system-dynamics-exploring-the-role-of-xbrl-in-an-environmental-reporting-system>

List of industrial publications and conferences

Publications

- 8) Alamillo, N., Cuesta, Y., Esteban, E., Fernandez, C., Flores, F., Mora, M. and Ortega M. (2017) “Big Data” e información empresarial (Big Data and Business Information). *Revista de la Asociación Española de Contabilidad y Administración de Empresas (AECA)*. ISBN: 978-84-16286-32-4 (<https://goo.gl/cRsNbQ>)

Two chapters of this book were written by Maria Mora, with revisions from other authors. The content was developed through project team discussions led by Bank of Spain.

- 9) Muñoz, F. F., Álvarez, J. L. L., Mora, M. and López, M. R. (2016). El nuevo Estado No Financiero y su verificación obligatoria: exigidos a partir de 2017. (The new Non-Financial State and its mandatory verification: required from 2017). *Revista contable*, (67), 18-21. (<https://goo.gl/Qx8FcZ>)

Several sections of this paper were written by Maria Mora, with revisions from other authors. The content was developed through project team discussions led by The Spanish Accounting and Business Administration Association (AECA).

- 10) Álvarez, J. L. L., Muñoz, F. F., Mora, M., and López, M. R. (2015). Nuevos horizontes de regulación sobre información no financiera información integrada, norma europea y propuesta de AECA. (New horizons of regulation on non-financial information integrated information, European standard and proposal of AECA). *Revista contable*, (36), 92-105. (http://is.aeca.es/wp-content/uploads/resources/articulo_AECA.pdf)

Several sections of this paper were written by Maria Mora, with revisions from other authors. The content was developed through project team discussions led by The Spanish Accounting and Business Administration Association.

- 11) Álvarez, J. L. L., Muñoz, F. F., Mora, M., and Lopez, M. R. (2013). La información integrada y el modelo propuesto por AECA. (The integrated information and the model proposed by AECA.). *International Journal of Accounting and Auditing*. Number 53.

Several sections of this paper were written by Maria Mora, with revisions from other authors. The content was developed through project team discussions led by The Spanish Accounting and Business Administration Association.

Speaker at conferences

- 12) Mora, M. (2017). Concept and use of enumerations 1.1: The CDP taxonomy. 23th Eurofiling Wokshop (European Central Bank, Frankfurt, Germany). June 2017

Work developed and presented by Maria Mora. Presentation available here: <http://eurofiling.info/2017/tutorials/enumerations/>

- 13) Mora, M. (2017). The EU Non-Financial Directive (2014/95/Eu) and its transposition in Spain. 19th XBRL Europe Day (European Central Bank, Frankfurt, Germany). June 2017

Work developed and presented by Maria Mora. Presentation available here: http://eurofiling.info/2017/wp-content/uploads/XEU_13_NFRD_MariaMora.pdf

- 14) Mora, M. (2016). Información Integrada y XBRL. Plataforma Integrated Suite. (*Integrated information and the Integrated Suite platform*). Risk Management, Interoperability and Information Standardization Conference (Business Registers, Madrid, Spain). December 2016

Work developed and presented by Maria Mora. Presentation and information available here: <http://www.xbrl.es/es/gestion-del-riesgo-interoperabilidad-y-estandarizacion-informativa/>

- 15) Mora, M. (2016). Envision climate-related reporting challenges. *Natural Capital Summit Conference* (Madrid, Spain). July 2016

Work developed and presented by Maria Mora. Presentation and information available here: <https://naturalcapitalsummit.wordpress.com/programa/>

- 16) Mora, M. (2016). Open data platform for integrated reporting in Spain. *22nd Eurofiling Conference*. (European Central Bank, Frankfurt, Germany). June 2016
Work developed and presented by Maria Mora. Presentation and information available here: <http://eurofiling.info/portal/xbml-week-frankfurt-2016-2/>
- 17) Mora, M. (2016). Financial Stability Board on Climate-related disclosure. *18th XBRL Europe Day Conference*. (European Central Bank, Frankfurt, Germany). June 2016
Work developed and presented by Maria Mora. Presentation and information available here: <http://eurofiling.info/portal/xbml-week-frankfurt-2016-2/>
- 18) Lizcano J. and Mora M. (2016). La vision del reporte integrado de AECA. (Integrated reporting vision by AECA). *Indra - Integrated Reporting Information Day*. May 2016
Work developed and presented by José Luis Lizcano and Maria Mora. Presentation and information available here: <http://is.aeca.es/aeca-participa-en-la-jornada-de-informacion-integrada-de-indra/>
- 19) Mora, M. (2015). Natural Capital Disclosure. *15th XBRL Europe Day Conference*. (Bank of Spain, Madrid, Spain)
Work developed and presented by Maria Mora. Presentation and information available here: <http://www.eurofiling.info/201506/index.shtml>
- 20) Mora, M. (2015). Standards convergence on Semantic Web: Linked Data. *21st Eurofiling Conference*. (Bank of Spain, Madrid, Spain)
Work developed and presented by Maria Mora. Presentation and information available here: <http://www.eurofiling.info/201506/index.shtml>
- 21) Mora, M. (2014). Climate Change Taxonomy Insight. *20th Eurofiling Conference*. (National Bank of Belgium, Belgium, Brussels)
Work developed and presented by Maria Mora. Presentation and information available here: <http://www.eurofiling.info/201411/index.shtml>
- 22) Mora, M. (2013). Enumerations. *18th Eurofiling Conference*. (CSSF Luxembourg, Luxembourg, Denmark)
Work developed and presented by Maria Mora. Presentation and information available here: <http://www.eurofiling.info/201312/index.shtml>

Activities and roles

- 1) Expert member of the International Financial Reporting Standards Board (February 2019-present)
- 2) Expert member of the ISO TC322 for Sustainable Finance (March 2019 – present)
- 3) Expert member of the ISO TC68 for Financial Services (March 2019 -present)
- 4) Chair of the Best Practice Board at **XBRL International** (March 2018 – present)
More information: <https://www.xbrl.org/the-consortium/get-involved/best-practices-board/>
- 5) Chair of the Academic Track at the **XBRL Week 2018** (January 2018 -present)
- 6) Principal Engineer on Artificial Intelligence and XBRL Expert at Fujitsu Laboratories of Europe (November 2017 – present).
- 7) Chair of the Academic Working Group at **XBRL Europe** (March 2017 – present).
- 8) Vice-Chair of the Best Practice Board at **XBRL International** (September 2017 – March 2018).
- 9) Member of the Best Practice Board at **XBRL International** (October 2016 – August 2017).
- 10) Coordinator of the Academic Track - **Eurofiling XBRL week** (October 2016 – December 2017).
- 11) Member of the **ICAEW** Natural Capital Accounting - Structured Data Task Force (INCA-SDTF). (May 2017-present).
- 12) Member of the Financial Reporting Lab working group led by the **Financial Reporting Council** (FRC). May 2017-present
- 13) Technical Manager at **CDP** (July 2013 - October 2017).
- 14) Technical leader of the Integrated Suite Working Group at **the Spanish Accounting and Business Administration Association (AECA)** (2012-present). More information here: <http://is.aeca.es/en/team/>
- 15) Member of the Spanish Accounting and Business Administration Association (**AECA**) (2006 -present).

Awards

- 1) Outstanding contributor in 2018 for chairing the Academic track at the XBRL Week and the work produced as the Chair of the Best Practice Board at the XBRL International, together with her long-standing active collaboration within the community. <http://eurofiling.info/portal/hall-of-fame/>

Chapter 1. Introduction

This chapter presents the research questions motivating this thesis, examines the academic and industrial background, and provides an overview of the document's structure and organisation.

1.1 Background and motivation

Environmental sustainability is defined by the United Nations [1] as responsible interaction with the environment to avoid depletion or degradation of natural resources and allow for long-term environmental quality.

According to a recent United Nations Environment Programme report on emissions, global greenhouse gas emissions have to peak by 2020 and decline rapidly thereafter to limit the increase in the global average temperature to no more than 1.5°C above pre-industrial levels. However, based on current policies and commitments, “global emissions are not even estimated to peak by 2030—let alone by 2020”. As a result, companies are under constant pressure from consumers, investors and governments to take responsibility for their impact on the environment. It seems clear that future business plans depend directly on a company's ability to be perceived as sustainable and to receive the approval of a broad set of stakeholders, comprised of shareholders, investors, employees, customers, suppliers, and the environment, mainly by means of an adequate reporting process to maximise stakeholder engagement and impact.

Currently, organisations implement different reporting processes. Where environmental information is a part of it, they normally take the form of sustainability reports, corporate social responsibility reports, annual reports, integrated reports and management reports disclosed under either a voluntary or mandatory basis. These are the mechanisms through which corporations share information about their economic, social, environmental and governance performance with their stakeholders to demonstrate the link between their strategy and its commitment to a sustainable global economy [2][3].

However, although this information has been generated over the past two decades, the actual impact of its disclosure is a key topic of discussion in both industry and academia. It is not clear that all types of reporting have the same impact on business and market performance. In the context of this study, it is considered that if businesses are both significant contributors to, and potential solution providers for, environmental problems, then reporting needs to be accurate and transparent to fulfil this role.

In exploring how to address this problem, attention is focussed on the interoperability concept initially defined by the IEEE as the *“ability of a system or a product to work with other systems or products without special effort on the part of the customer. Interoperability is made possible by the implementation of standards”*. This concept has evolved, and it is understood today as the ability to exchange and use information given a heterogeneous landscape of organisations and information [4], beyond the merely technical layer. Interoperability is relevant due to how the corporate reporting ecosystem represents a complex scenario of information (Figure 1). It is composed of data covering a variety of financial and non-financial topics from different organisations, data formats (video, tweets, spreadsheet, XBRL, HTML, PDF, sensors, etc.) and reporting frequencies (annually, quarterly, real time). In the past, the main focus of this data was to satisfy the information needs of shareholders. Now, it is necessary to evolve from shareholders to a broad set of stakeholders, which includes almost all global public opinion; as a result, the boundary between shareholders and stakeholders is becoming blurred. For many years, leading companies have found that the integration of social, environmental and environmental objectives into their broader operational and financial missions assists regulatory compliance, and can also become a basis for developing unique competitive advantages. This allows them to respond effectively and proactively to the increasing social and environmental responsibility demands of customers, insurance companies, green investors, ethical trusts and innovative competitors [5][6]. This alignment with the stakeholder means both financial and non-financial issues must be considered in corporate decision-making [7].

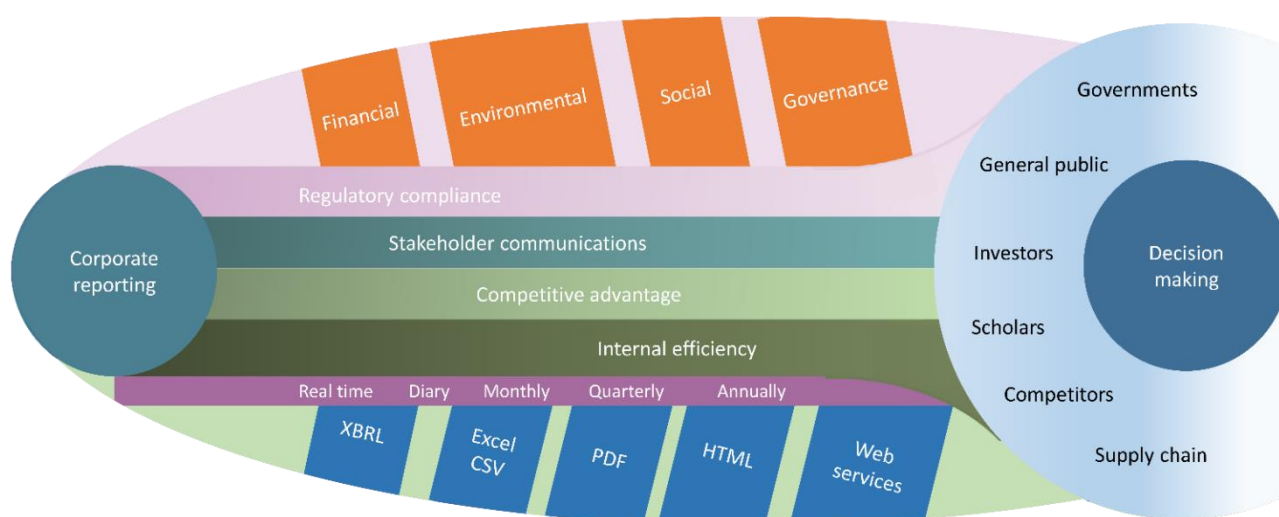


Figure 1. Corporate reporting ecosystem

This information ecosystem brings potential opportunities to increase the impact of environmental information on the decision-making processes of companies and their stakeholders. However, this ecosystem poses unique challenges, such as:

- How to provide better information? First, to promote more effective environmental behaviours, and second, to enable people to take a more informed approach to their environmental decisions. This would facilitate stakeholder engagement and communication.
- How can companies and stakeholders improve their decision-making processes? This involves a company's internal processes at the strategic, management and operational levels to answer questions such as: What needs to be done to enable investors to better analyse companies' risks and opportunities considering their environmental implications? How can the general public easily judge a company based on its environmental and societal impact?

This thesis can inspire innovations in IS solutions to support the production, distribution and consumption of corporate information, taking into consideration the different problems and challenges summarised in Figure 2.

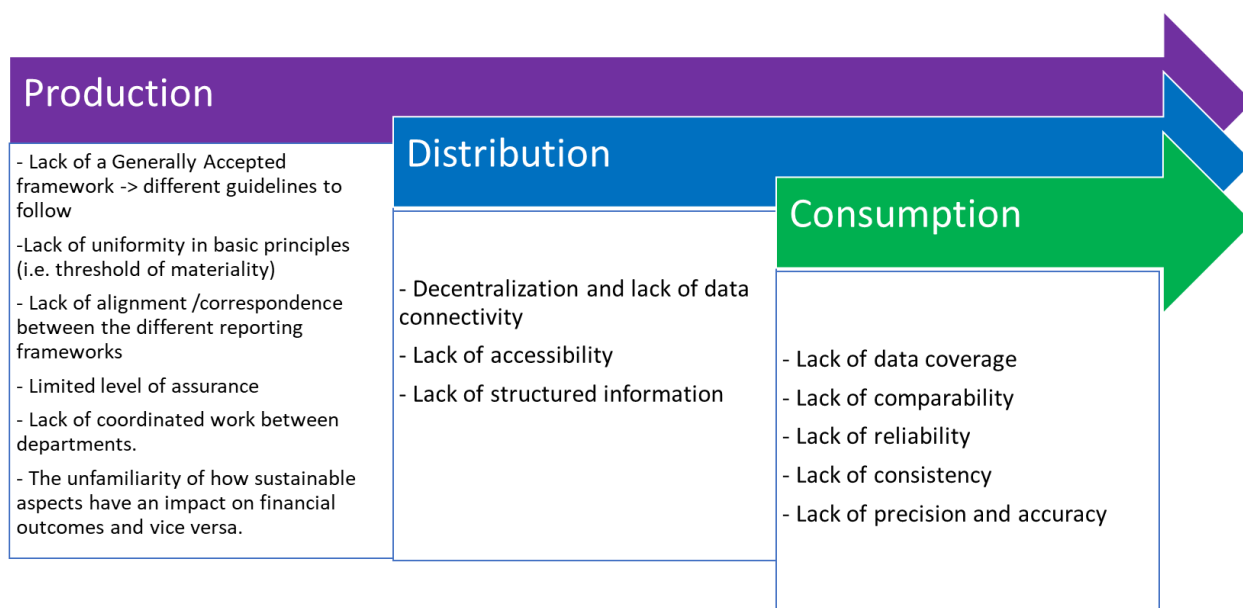


Figure 2. Problems and challenges in corporate reporting processes

To clarify that this thesis is part of an Engineering Doctorate (EngD) programme instead of a traditional PhD. The main difference lies in the fact that the result of this research has to demonstrate a contribution not only within academia but also in industry. To achieve the latter, the market and application into the industrial sponsor of this work “The Carbon Disclosure Project (CDP)” was considered during the course of the study .

1.2 About the industrial sponsor of this EngD

Formed in 2002, CDP is a leading organisation in the environmental reporting field and a global standard setter. Looking at CDP as an agent within a network of actors (Figure 3), it has collaborated with companies, investors, NGOs, governments and cities from across the globe, helping them to disclose, measure, manage and share their environmental impact, awareness and actions. CDP supports organisations by helping them be more accurately informed about their environmental performances and consequently better able to communicate the information and related strategies to their stakeholders. CDP works with influential market

players, including 827 institutional investors with assets of US\$100 trillion. It holds the largest global collection of information on primary climate change, water and forest risk commodities, and puts these insights at the heart of strategic business, investment and policy decisions. 5,600 organisations and 500 cities now use CDP to disclose vital environmental information to investors and major purchasers. Currently, CDP maintains collaboration lines to align areas of reporting with other disclosure initiatives, such as the Global Reporting Initiative (GRI) and the International Integrated Reporting Framework (IIRC) as described in more detail in Chapter 3. The main goal of this collaboration is to enable users and reporting firms to refer to the same data points through different reporting channels. It represents a significant step towards the global standardisation of environmental reporting.

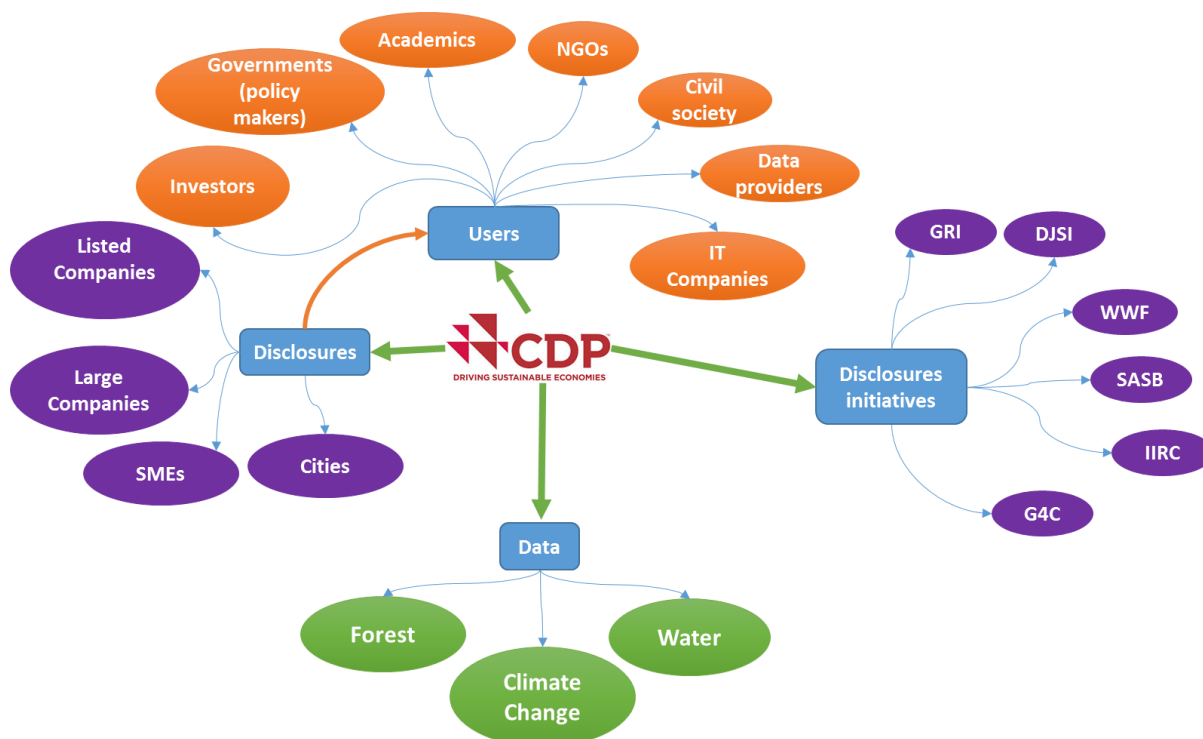


Figure 3. System map: Exploring CDP as a network of actors

The reason for this project is that although CDP offers one of the most comprehensive corporate environmental datasets in the world, and is useful to a number of actors (Figure 2), it is not yet clear what the real use and impact of this data is on decision makers. Questions arise regarding the place of CDP in the market as an environmental data provider in the corporate reporting ecosystem and its impact on decision-makers, and this thesis is focussed

on how Information Systems (IS) can bring solutions and new opportunities to drive the business transformation.

1.3 Purpose of this research project

This research project aims to provide additional insights on the role of Information Systems in environmental sustainability. My attention is focussed on CDP as a useful dataset and its current role as a participatory platform, promoting the disclosure and use of environmental data to a variety of stakeholders. This work explores the value provided by insight into environmental information and how new data technologies and data science techniques can be implemented on top of it using real case studies, to present ways to the market and general society to make CDP data more relevant for decision-making.

From an academic perspective, the project is linked to Accounting Information Systems and socio-technological disciplines, and applies system engineering techniques with a dual aim. First, to develop a good understanding of stakeholder requirements on CDP information systems. Second, to make technical contributions in the area of reporting standards, semantic technologies and data science techniques that will support environmental reporting challenges and bring new opportunities to increase its impact on stakeholders. This work is relevant for academia and software engineers to build IS for environmental sustainability, especially given the research interest in open government datasets and Big Data tools that are being used by some research groups in the context of corporate reporting and Accounting Information Systems.

The industrial contribution of this work is the strengthening of CDP's approach to data, through an efficient use of data and tools based on data standards, semantic technologies and data science techniques. This work is a generator of ideas for CDP that drives the company's technology strategy, and educates and engages practitioners in innovative technology developments and trends. This work helps CDP achieve the next stage of its growth and create impact.

1.4 Thesis structure

This thesis has four main goals: (1) Identify the problem that I am trying to solve, taking into account previous academic efforts and the reality of the market. (2) Introduce my visions exploring current activities, future projections, opportunities and continuing challenges. (3) Propose solutions to tackle the challenges identified and to foster the opportunities explored previously and (4) create impact. The structure of the thesis is introduced below and presented in Figure 4.

- **Chapter 2** describes de different methodologies applied during the thesis.
- **Chapter 3** describes the context of this research. It identifies a set of potential research goals of interest in academia and industry about CDP and the impact of its disclosure, corporate reporting needs and requirements, the use of Information Systems and technologies that can bring solutions in this domain.
- **Chapter 4** presents the relevant literature related to this research, where the key concepts, theories and research gaps about the role of Information Systems in environmental sustainability and the effectiveness of disclosure are explored. The main contribution of this chapter are:
 - Identifying the reasons why companies disclose environmental sustainability information, how they disclose such information, and its current impact.
 - The identification of several reporting challenges, such as the unfamiliarity of how sustainable aspects have an impact on financial outcomes and vice-versa, and the lack of comparability and consistency. A clear strategy to promote solutions which support them is also lacking.
 - The initial presentation of key technologies that can potentially provide solutions to some of the highlighted challenges, especially on data standardisation, data connectivity and data analysis.
- In **Chapter 5**, it is postulated that better quality data and Information Systems (IS) are critical to ensuring business transformation. However, evidence from previous chapters demonstrates that IS are not entirely prepared to cover stakeholder demands. In this chapter, the results of interviews with 21 CDP members are

presented. These interviews included representatives from companies, investors, governments, academia, NGOs and software/data providers. This chapter contributes to identifying stakeholder requirements on CDP information systems on four dimensions: drivers for participating in CDP; barriers that CDP stakeholders face either regarding reporting performance or using the data; prescriptive solutions to improve adoption of CDP; the impact of CDP participation and use of its data. The results show clear demands for access to a larger and more diverse environmental dataset as well as innovative IS solutions to support data quality, interoperability, accessibility, analysis and visualisation.

- **Chapter 6** is driven by the results of Chapter 5 about the critical role that interoperability plays on the impact that decisions made by businesses, corporations and institutions have on the environment. This chapter explores the role of “interoperability” as a critical component of business transformation and decision-making in complex information scenarios. The outcomes offer some insights on new opportunities to build better solutions to overcome sustainability challenges.
- **Chapter 7** envisions the role of reporting technologies in supporting climate reporting and its standardisation. Specifically eXtensible Business Reporting Language (XBRL) is presented as a potential mechanism to overcome existing corporate reporting challenges and maximise the future development of a sustainable economy. Potentially, this study connects with the ideas of better data standardisation and technical and semantic interoperability discussed in Chapters 5 and 6. This chapter represents a significant industrial contribution to influencing the practices of XBRL within CDP and its market. As a result, I produced (1) a ground-breaking new report, focusing on exactly how XBRL can help overcome climate-related reporting challenges, and (2) the implementation of the 2016 CDP-Climate change taxonomy, which is the digital representation of CDP data in XBRL format. This taxonomy encourages the utilisation of Climate Change data in business strategies and operations, and its introduction to the public. It represents the starting point for organisations to create value using digital reporting standards and help incorporate Climate Change data into decision-making. This work contributed to:

- Improving data quality and data accessibility: standardising Climate Change data and business rules in a digital open format.
 - Connecting environmental and financial information models: using the same reporting technology adopted by financial information markets worldwide (e.g. IFRS, USGAAP, UKGAAP, Spanish GAAP, EDINET, etc). XBRL is already required around the world by regulators and supervisory agencies.
 - Improving the consistency of environmental data across other sustainability disclosure frameworks and reducing reporting burdens: aligning data contexts across other sustainability frameworks which use XBRL, such as GRI (Global Reporting Initiative) and DJSI (Dow Jones Sustainability Index).
- **Chapter 8** presents a solution to enable a major integration of corporate financial and environmental data using XBRL and Linked Data principles. This solution addresses the properties envisioned in Chapter 6 and which Chapter 5 identified as relevant to stakeholders' needs (*"CDP should make accessible its data linked to other datasets"*). This study presents a solution which demonstrate how Linked Data technologies can be used to build a bridge between financial and environmental data published in XBRL format and other Open Data silos, such as DBPedia (the structured database of Wikipedia), implementing a real solution that contribute to the benefits on better data contextualisation, validation and data analysis. These solutions should be considered by initiatives like CDP at the publication stage of its data. This work was presented to experts at the 14th European Semantic Web Conference (ESWC).
- **Chapter 9** proposes a solution using text-mining analysis to determine alignments between different corporate reporting frameworks that share common data, in this case the GRI and CDP reporting frameworks. The misalignment between different corporate reporting frameworks is one the difficulties identified during the stakeholders' needs study in Chapter 5, and so this chapter proposes solutions for automatically assessing levels of correspondence between frameworks which represents a relevant contribution.

The thesis concludes with **Chapter 10**, which summarises the most significant impact and conclusions from my work. Suggestions for further research is also discussed.

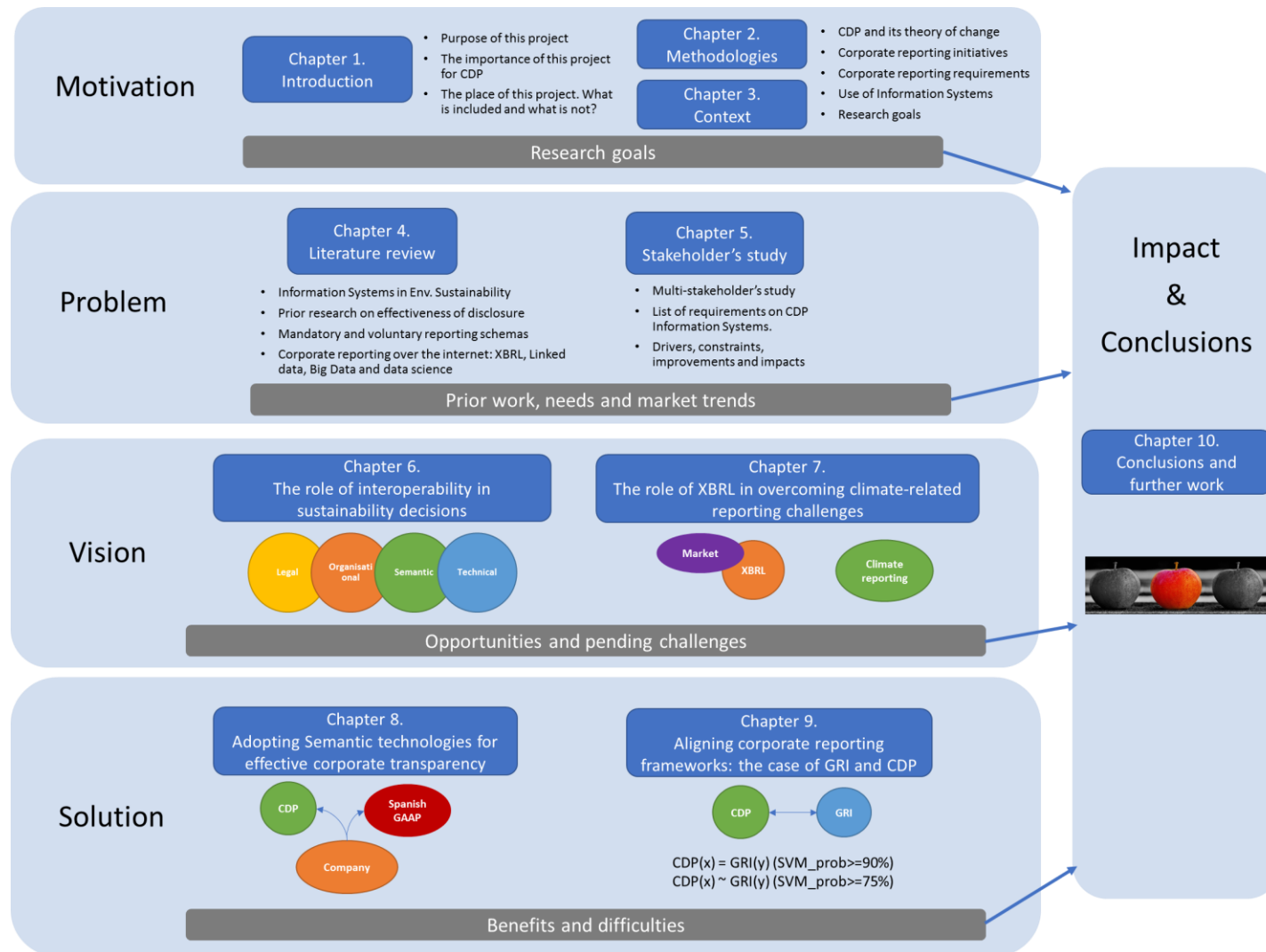


Figure 4. Thesis structure

Chapter 2. Methodologies

This chapter describes the methods applied in this thesis.

2.1 Qualitative research methodology

In chapters 5 and 6, we analyse and explain organizational behaviours relating to environmental reporting initiatives. For that purpose, we conducted interviews with six different categories of stakeholders: companies, investors, governments, academics, software/data providers and NGOs. The participants are CDP members from Europe, the United States of America and South America. All of them have a thorough understanding of the CDP data and tools as users and informants. The sampling method selected was the non-probability sampling method called “Convenience sampling” because participants were selected via CDP Directors based on participant’s availability and willingness to take part.

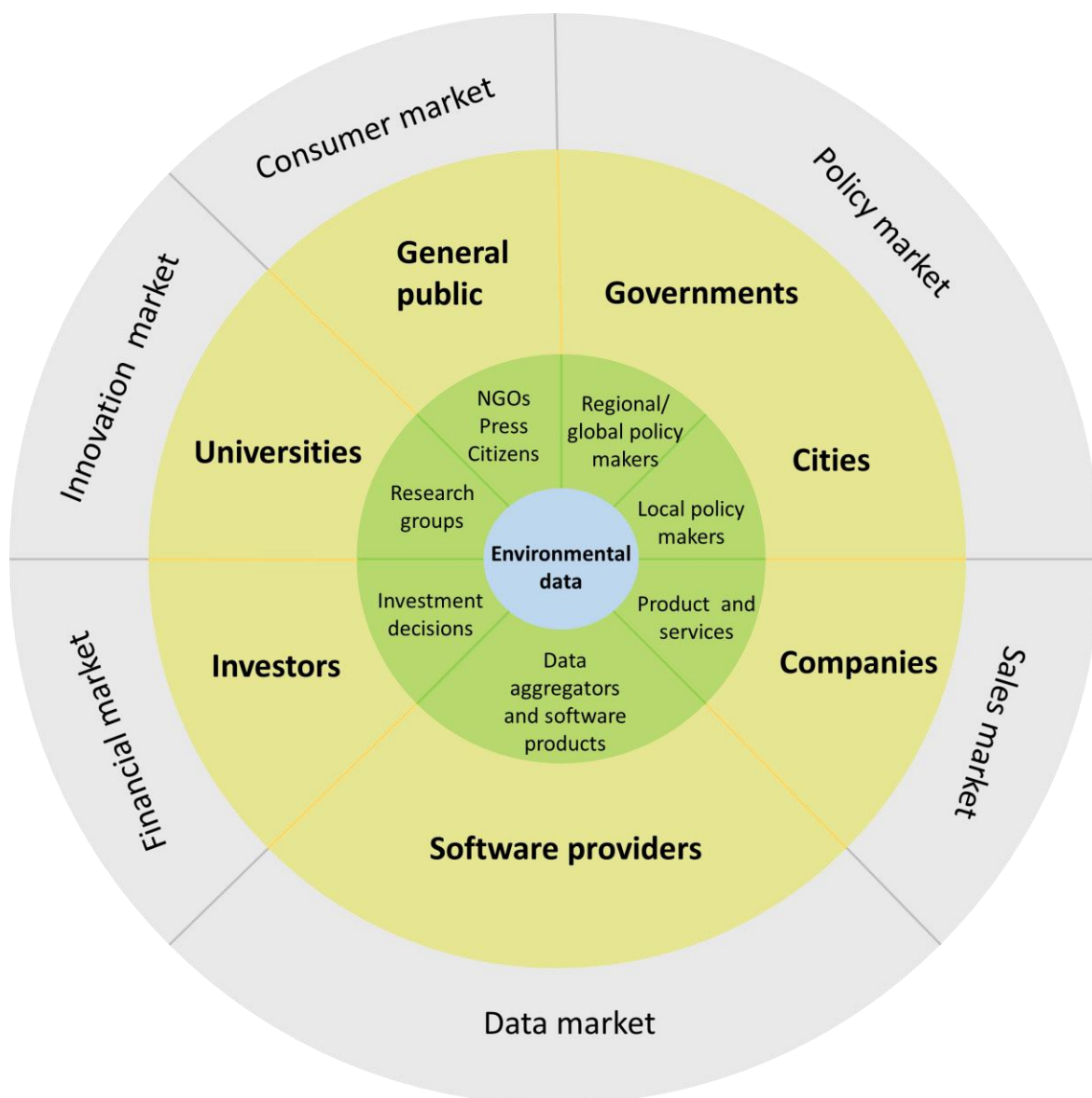


Figure 5. Stakeholders and market map

These studies required the evaluation of a whole spectrum of CDP stakeholders. Semi-structured interviews with open questions were selected as the research method, allowing engagement with a set of experts in a process designed to generate individual ideas and achieve synthesis and clarity on the research questions [8]. Moreover, interviews provide the ability to clarify responses and get supplementary information if required [9][10]. The interviews were conducted in Spanish and English with qualified experts who have a professional understanding of environmental and corporate data. Thirty-five qualified data analysts and senior manager participants, divided into the seven stakeholders groups presented in Figure 5, were invited to participate through an email from CDP. However, only

21 members agreed to be interviewed, the majority non-native English speakers. Information about the participants and the groups to which they belong is presented in Table 1. To maintain the anonymity of the participants, only the number of organisations per group has been identified. The companies that agreed to participate belong to different industries, (Oil and Gas, Telecommunication, Chemical and Explosive). The investors were international asset management companies. Government groups were represented by global organisations and local authorities. Meanwhile, academics were research group experts in a diverse number of business fields (policy, environmental and financial risks and performance and accountability). The NGOs were composed of non-profit organisations that work internationally with businesses and governments. Software and data providers were represented by companies that develop software for corporate reporting and data aggregators.

Stakeholder group	Participants
Companies	5
Investors	4
Governments	2
Academia	4
General public (NGOs)	2
Software/Data providers	4
Total	21

Table 1. Distribution of participants per stakeholder group

The interviewees received the questionnaire and a brief introduction to the study about one week in advance of their interviews. Face-to-face interviews were conducted with 58% of the participants, and the remainder interviewed via video conferences over the course of six months. The interviews were recorded and then transcribed. After the transcription, some additional questions were asked of certain participants to clarify their responses [11]. Codes were developed following the recommendations of [12][13][14] to identify emerging patterns and themes, and to identify relevant categories for analysis. To improve the reliability of the results after three rounds of coding, we developed the categorisation of drivers, constraints, improvements and impacts for chapter 5 and for chapter 6 the interoperability frameworks (for more information refers to chapter 6). Answers were analysed against the research

questions and the conceptual frameworks. To support the coding and analysis, we used the qualitative data analysis and research software called ATLAS.ti¹.

2.2 Taxonomy design methodology

In chapter 7, we developed a solution using XBRL technology to standardise environmental data, given as a result an XBRL taxonomy for environmental data. For that, we used a phased approach (Figure 6) to take along the development path.

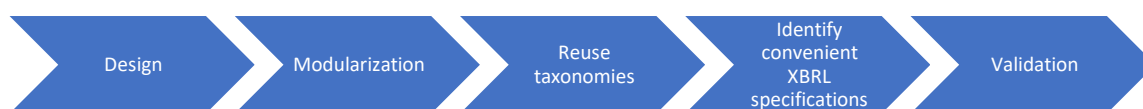


Figure 6. Taxonomy Design methodology

At the design phase, we built a data dictionary with the goal to reflect the value of the information contained. At this stage we avoided to constraint at early stages the value of the information due to technical restrictions in the XBRL Specification. Once the domain of the information was defined on spreadsheets in excel, we checked with the CDP Technical Director. Once, all the elements were identified, defined and checked, we decided the level of modularization to break the taxonomy into different parts, this is important to facilitate its applicability in different sectors (such as Oil and Gas, Finance..) and also for the maintenance of the XBRL taxonomy in the medium and long-term given the CDP reporting guidance is updated every year with new data points and existing ones are removed. We also evaluated the possibility to reuse existing taxonomies, as well as, XBRL specifications. As a result, the taxonomy was modularized in different parts to 1) facilitate the maintenance of the schema in the medium and long-term and 2) open the possibility to address other specific sectors beyond Oil and Gas. Formulas², Extensible Enumeration³ and Dimension⁴ specifications were

¹ Atlas.ti Qualitative Data analysis: <http://atlasti.com/>

² Formula specification: <https://specifications.xbrl.org/work-product-index-formula-formula-1.0.html>

³ Extensible Enumeration specification: <https://www.xbrl.org/Specification/extensible-enumerations-2.0/CR-2019-01-09/extensible-enumerations-2.0-CR-2019-01-09.html>

⁴ Dimensions Specification: <http://www.xbrl.org/specification/dimensions/rec-2012-01-25/dimensions-rec-2006-09-18+corrected-errata-2012-01-25-clean.html>

selected to better represent the validation rules and qualitative information. There was not necessary to reuse existing taxonomies.

Once the taxonomy was built, we validated against XBRL specification and related filing rules.

2.3 Machine Learning methodology

In chapter 9, a Machine Learning methodology was developed to facilitate the identification of common data points (such as risk and opportunities in climate , policy influence, etc) from different reporting frameworks. To demonstrate the usefulness of our proposal, we tested our method using CDP and Global Reporting Initiative (GRI) datasets. We proposed the approach of “learning from massive data” [15] by analysing texts from GRI and CDP reports. These analyses were carried out by means of Natural Language Processing (NLP) methods, specifically supervised learning for classification tasks [16].

In Figure 7 we illustrate the methodology we followed:

- (1) to generate and validate the classification model using **training** data (source of learning material), which in our case is the CDP responses classified by questions.
- (2) to **predict** the classification of a new dataset applying the new model. In this instance it means classifying the GRI reports into CDP questions to get an accurate measure.

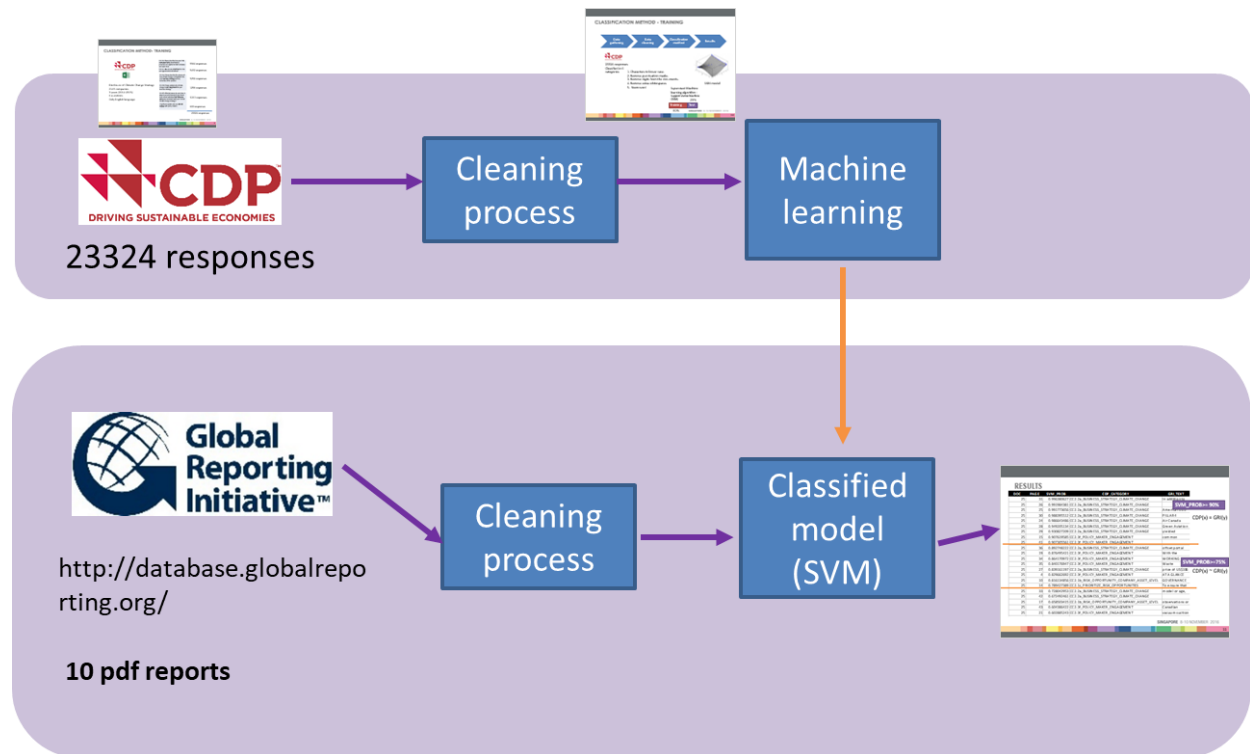


Figure 7. Text classification methodology

Chapter 3. Context

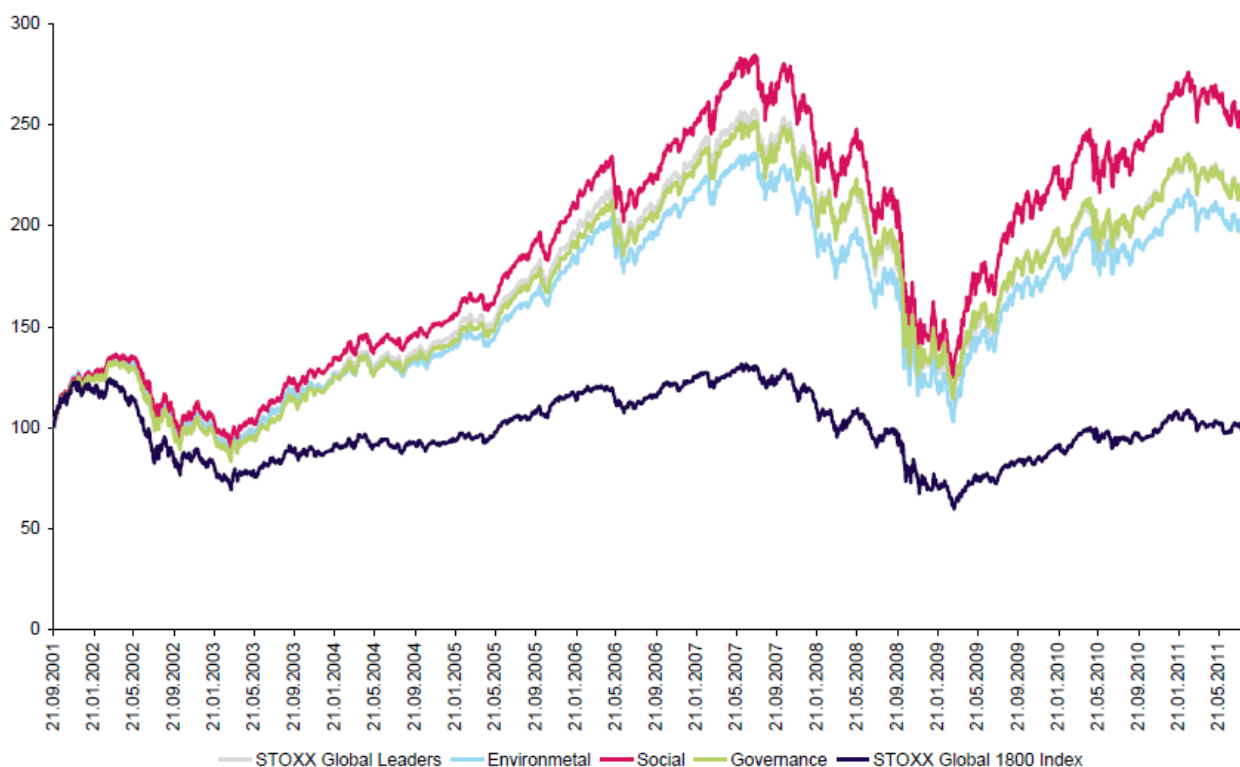
This chapter presents a set of research questions of interest to academia and industry about CDP and the impact of disclosure, corporate reporting needs and requirements. It also provides background about the use of Information Systems (IS) and technologies that can offer solutions in this domain.

3.1 Background

More than 12,000 public companies worldwide have recently issued reports on various aspects of their environmental sustainability mostly in response to investors' demands and regulators' requirements for this information [20]. Global stock exchanges are either requiring or recommending listed companies to report their environmental sustainability in addition to their financial performance. Since 2015, listed companies on the Hong Kong Stock Exchange have required disclosure of sustainability information. About 6,000 large European companies are now required to disclose information about their environmental, social and governance (ESG) sustainability performance in the 2017 reporting year [21]. More than 8,000 public companies across the globe issued a stand-alone sustainability report in 2015, compared with less than 500 companies in 2005 [20]. Today there is empirical evidence in several industrial sectors that sustainability reporting has become of competitive relevance and strategic importance [22][23].

With the 2008 global financial crisis, there was an erosion of trust in states and corporations, and companies were called on to play a role in restoring it. As a result, environmental dimension was developed to help produce more stable and predictable economic systems and markets. Previous research has found that environmental improvements can help corporations to save resources, and to increase sales by means of product enhancements and quality. However, due to their reporting practices and its positive impact on reputation, the companies that more rigorously adopt a responsible strategy can benefit from significantly better stock market behaviour [24][25]. What seems clear is that the one vehicle by which firms benefit from improving their results is reporting. According to STOXX [26], it is possible to observe how companies listed in a "sustainability index" present better historical behaviour in the stock market price of their shares than a general sample of quoted

corporations (Figure 8). Research are showing that companies which do not integrate environmental factors into their business strategy put their long-term competitiveness at risk. The consideration of environmental factors leads companies to have a stronger long-term position.



Graphic: Outperformance of broad market indices (data, as of August 23, 2011)

Figure 8. The STOXX Sustainability index against a general stock market index

Is it that companies following environmental sustainability strategies can generate greater value for their shareholders? Alternatively, is it the other way around: the best-performing companies have enough resources to take care of their stakeholders once the shareholders are satisfied? Is it X that generates Y or vice versa? This is still a controversial topic in the academic and professional literature. However, some relevant authors argue in favour of the first option [27][28][29]. For many years, leading companies have found that the integration of environmental objectives into their broader operational and financial missions can assist regulatory compliance.

When dealing with environmental reporting, a key challenge arises: the need for a standard to guide companies in preparing and disclosing such information. Following the development of a certain number of consortia and guidelines, companies have dramatically increased the amount of information disclosed by reporting not only financial but also environmental issues in longer and more complex annual reports, or by providing sustainability reports along with the traditional financial data.

More information is not necessarily synonymous with better or more relevant information for decision making. The diagnosis could be as follows: at the moment, large amounts of information are generated, certainly in excess. Such information is expensive to develop, especially if we expect to meet the increasing regulatory demands derived from the economic crisis.

Today, corporate reporting continues to grow in a disjointed manner. Financial reports have significantly expanded. At the same time, new information is being developed for different purposes in a disconnected manner. Annual reports, sustainability reports, corporate governance reports and other documents on company management are not linked. They do not reference each other and they are often developed using divergent and/or overlapping reporting definitions. Each of these reports is expressed in very different terms and styles, with varying degrees of complexity depending on their target audience and the framework that governs their construction.

Ease of comparability, ease of consumption and the levels of reliance and trust that a user may ascribe to these different reports are uneven. This creates significant difficulties in the processing and analysis of information by users, especially investors and analysts.

In short, the current situation of corporate reporting could be described in the following way: it presents a large volume of unrelated data that is not necessarily relevant, is expensive to produce, is difficult to handle and is of little strategic significance (or reduced usefulness) for future-orientated decision making given its eminently historical nature. To summarise, the current deficiencies of corporate reporting are:

- The variety of formats provided (PDFs, HTML, XBRL, Excel, etc.).

- The lack of quantitative indicators.
- The misalignment in the basic principles (e.g. thresholds of materiality that relate to the significance, relevance, reliability of the information for the performance of the firm), especially because of the variety of guidelines available to create such reports.
- The lack of reliability, due to weak auditing practices regarding the information provided and the applied verification processes.

Given these corporate reporting requirements and technical problems, we found that Information Systems, digital reporting standards, such as XBRL (eXtensible Business Reporting Language)(Section 2.4) and Big Data techniques, semantic technologies (section 2.5) provide powerful tools to turn isolated pieces of data and reports into valuable information ready for decision-makers. Chapters 7, 8 and 9 make primarily technical and academic contributions using these tools in the context of corporate reporting.

3.2 Voluntary and mandatory reporting schemas

The disclosure of environmental sustainability is happening through a multitude of voluntary and mandatory reporting initiatives. Annex A presents a set of reporting initiatives that directly or indirectly relate to this research. We describe their requirements and the most relevant reporting features, including whether the disclosure is a government requirement, guidelines or reporting system; country and disclosers affected; the type of information disclosed; whether the initiative is mandatory or voluntary; whether this information usually comes in a type of report (annual report, CSR reports, Integrated report); the communication channel; data format; whether the information is publicly available or not; external frameworks referenced; and disclosure location.

As a general conclusion from Annex A, we found common domains of information disclosed under different reports and data formats. In some cases, this information is centralised by an organisation and freely accessible, or is available at a cost, as is the case of CDP. The majority of these initiatives reference multiple frameworks that differ in the level of detail proposed. For example, looking at the European Directive, it references CDP and IIRC as valid frameworks to comply with the Directive, but it does not specify how and where each of these

frameworks may be used to comply with the Directive. This fact complicates matters for disclosers to understand how and when to use each framework to support their disclosures. For example, CDP is a data-driven framework while IIRC is a principles-driven framework, which references CDP as part of its natural capital work. In general, the different reporting initiatives presented demonstrate that there is a lack of a unique Generally Accepted Accounting Principle (GAAP) for the disclosure of sustainability information. It also reveals that there is a lack of understanding about how these different frameworks should work together to maximise the value of the reported information and create a real impact on the final users.

3.3 Corporate reporting and market behaviour

When dealing with non-financial reporting, a key challenge is the need for standards to guide companies in preparing and disclosing such reports.

Since 1991, there have been several projects aiming to respond to the need for a standard to guide companies in preparing and disclosing sustainability performance. In that year the AICPA (American Institute of Certified Public Accountants) established the Special Committee on Financial Reporting, known as the Jenkins Committee. This discussion forum was set up given the growing demands for an improved corporate reporting model able to analyse users' increasing demand for business information (focusing on investors and lenders), and to develop the content of company business reporting to accommodate users' needs. One of the most interesting attempts to improve accounting information comes from the document entitled "Improving Business Reporting – A Customer Focus"; commonly referred to as the Jenkins Report, it was issued in December 1994. The motivation of the Jenkins Report was to address the general dissatisfaction with the model of financial information.

Nearly a decade later, in January 2003, the American Institute of Certified Public Accountants established the Special Committee on Enhanced Business Reporting to take action over initiatives that had fallen into oblivion, such as the Jenkins Report. The Committee concluded its work in 2005, having brought together a consortium of investors, creditors, regulators, managers and other stakeholders to improve the quality and transparency of the information used for business decision making. Thus was born the Enhanced Business Reporting

Consortium (EBRC). The Enhanced Business Reporting Framework was published in October 2005 and was intended to promote greater transparency regarding the strategy and performance of businesses. It was based on the materials used to elaborate the value reporting model of PricewaterhouseCoopers, which researched the types of information used in 16 industries in 14 countries both to manage an organisation (information from the manager) and to assess an organisation for purposes of investment. This framework organises the disclosure of additional information not currently covered by the Generally Accepted Accounting Principles (GAAPs). The EBR Framework recommends companies to disclose information on corporate responsibility, from both its main perspectives: respect and protection of the natural environment, and commitment to social, ethical and charitable principles [30]. In addition, it stresses the importance of information technologies and suggests that companies should explain how they ensure that their technologies are operating as intended and how the integrity and reliability of information are assured. This reporting framework also contains disclosure items related to the analysis of the environment and the strategy of the company, combining historical and prospective reporting items. This conceptual framework contains four building blocks on which to structure business reports: business scenario, strategy, resources/processes and performance.

Concerning private organisations that have also had a significant influence on enterprises, we will mainly refer to AccountAbility, the Global Reporting Initiative (GRI), the International Integrated Reporting Council (IIRC) and the Carbon Disclosure Project (CDP). AccountAbility is a global non-profit entity, the purpose of which is to promote innovations in accountability that foster sustainable development. The network engages with businesses, governments and civil society organisations to advance responsible practices in business and management through the cooperation of public and private institutions.

In 1997, an international organisation based in Amsterdam brought together numerous agencies, associations and enterprises to launch the GRI. The GRI is comprised of an extensive network of stakeholders, including business organizations, NGOs, universities, etc., which, structured through thematic and industry working groups, develops guidelines to promote a conceptual framework allowing the continuous improvement of sustainability reporting. This standard is based on a conceptual framework that includes a set of principles and indicators that organisations must use to measure and report on the economic, social and environmental areas of their performance (triple bottom line). At present, over 10,300 organisations around the world organise the dissemination of information based on the GRI guidelines summarising the disclosure under the tenets of GRI [31].

With the aim of promoting better reporting practices for sustainability, the International Integrated Reporting Council (IIRC)⁵ was created in 2009. It is a global coalition of regulators, investors, companies, standard setters, the accounting profession and NGOs. The main idea behind integrated reporting is to create a high-level qualitative guideline on how reporting should be properly transmitted to the audience regarding the ability of the firm to create long-term value, combining performance indicators with data on corporate strategy and prospects, in addition to other aspects. The inception of the IIRC is one of the most visible initiatives worldwide promoting better reporting practices. The appearance of IIRC into the current reporting scene would appear to improve the quantity and quality of reports from the engaged firms, but it does not solve the pre-existing heterogeneity and lack of efficiency.

In 2002, the Carbon Disclosure Project (CDP), an international non-profit organisation, was launched as a global initiative to drive corporations to measure, disclose and manage their environmental risks and reduce their carbon emissions. The CDP is working with several standards such as the GHG Protocol, GRI, and IIRC⁶ to revolutionise the way in which the world's most powerful organisations report their climate change information.

⁵ IIRC: <https://integratedreporting.org/>

⁶ GHG Protocol: <http://www.ghgprotocol.org/>

Other standards that can be mentioned are the United Nations Intergovernmental Working Group of Experts on International Standards of Accounting and Reporting (ISAR) ⁷, which provides voluntary technical guidance on eco-efficiency indicators, corporate responsibility reporting and corporate governance disclosure, and the ISO 14000 environmental management standard. All these initiatives coexist with pending challenges on the harmonisation of regional or national requirements.

From an institutional point of view, three non-binding standards that have also had great influence on non-financial reporting are the United Nations Global Compact, the Conventions of the International Labour Organization⁸ and the OECD guidelines ⁹. The European Union has also issued several statements and recommendations on CSR and sustainable development, such as the European Strategy for Sustainable Development and the Green Paper on Promoting a European Framework for CSR. The most recent one is the European Directive 2014/95/EU for non-financial reporting which, in 2017, applies to public interest companies with more than 500 employees, representing approximately 6000 large companies in Europe. This Directive defines a set of information to be disclosed on social, environmental and risks matters, and suggests a set of voluntary reporting frameworks to enable compliance [21].

Following the development of such a number of consortia and guidelines, companies have dramatically increased the amount of information disclosed by reporting not only financial but also environmental and social issues in longer and more complex annual reports or by providing sustainability reports along with the traditional financial data. The main research question in this area is whether, after all these efforts, shareholders in particular and stakeholders in general are able to gain full advantage of this richer reporting environment and to which extent they significantly change market behaviour in a turbulent environment. Eleftheriadis and Anagnostopoulou [32], for instance, find a positive relationship between climate change and some firms' factors, but none concerning relevant aspects such as profitability or leverage. The effect of environmental reporting on both accounting and market variables is still a matter of active international research.

⁷ International Standards of Accounting and Reporting (ISAR): <https://isar.unctad.org/>

⁸ International Labour Organization (ILO): <https://www.ilo.org/>

⁹ OECD: <https://www.oecd.org/>

3.4 Use of information systems

All reporting needs to consider the significant technological developments that are changing the way that data are used, presented and analysed.

To enhance the threshold of environmental data utilisation, an organisation must recognise that it should further leverage available technologies. By doing so, they can create more value for themselves, reporting organisations and society.

One of the promising domains that is explored in this thesis is how to create impact with environmental data through reporting standards, procedures, methods and tools for assessing decisions across its disclosure information. Investors, governments, individual organisations and cross-sector reporting initiatives have approached corporate reporting initiatives, such as CDP, GRI and IIRC, with this interest in mind.

Significant challenges identified include:

- The lack of easily accessible corporate data to quickly and accurately inform management about material issues, which are pertinent during decision-making processes.
- Unfamiliarity with how sustainable aspects have an impact on financial outcomes and vice versa.
- Inadequate levels of integration of financial and non-financial information within the internal performance, strategy and operational frameworks of an organisation.
- A dearth of consistency, comparability, reliability and clarity of climate change information emerging from organisations globally. Standardisation and mainstreaming of disclosures need to be facilitated.

For these reasons, there have been several independent initiatives to try to overcome the challenges motivated by this enormous increase of corporate non-financial information available online, vaguely supported by the described succession of guidelines and frameworks. The AECA's project is one of these initiatives.

3.5 Corporate reporting initiatives

The AECA (Spanish Accounting and Business Administration Association, or *Asociación Española de Contabilidad y Administración de Empresas*) is a non-profit scientific and professional entity, which was founded in 1979 with the purpose of combining efforts to develop a doctrinal body on accounting and business administration. Through the issuance of generally accepted principles, rules and best practices following international trends, the AECA facilitated the introduction of accounting standards in Spain in the 1980s and 1990s. Thanks to this work and to the introduction of the AECA's Accounting Principles in the regulatory frameworks implemented later in Spain, professional accounting and auditing practice in Spain has reached international standards. The AECA has been leading various international theoretical and empirical studies on financial and non-financial reporting, transparency and corporate and public governance. Traditionally, the AECA researches business reality in order to illustrate and reinforce the validity of its studies.

One of the most recent projects promoted by the AECA is the so-called Integrated Scoreboard for Financial, Environmental, Social and Corporate Governance information (IS-FESG), for which the AECA provides technological support for the generation, transmission and processing of integrated reports on the strategic activities and situation of companies by means of the use of the Integrated Scoreboard of directly comparable indicators [33]. The use of digital standards like XML (eXtensible Markup Language) is intended to promote comparability between companies, to increase corporate transparency at the international level. In accordance with the AECA's nature, the Integrated Scoreboard (IS-FESG) is a royalty-free, open specification to describe business behaviour through a set of indicators available for public and private companies and other organizations, regardless of their national jurisdiction.

Solutions like digital standards in this context make information easy to find and access, providing the ability to locate individual data points and documents at a company or industry level. It could also go beyond the current system by embedding the context at the level of individual pieces of data or disclosures. For those purposes, digital standards and semantic technologies, such as XBRL (eXtensible Business Reporting Language), which is an open digital standard for exchanging business information based on XML, and Linked data appear as

potential solutions for corporate reporting problems. Chapters 7 and 8 are dedicated to exploring these solutions.

One of the most significant contributions of this research has been to make CDP recognise open reporting standards such as XBRL to promote a better standardisation and structured data of environmental information from companies and a better harmonization with financial reporting practices. To help accomplish this, further collaboration with users of corporate data was necessary. We identified what specific additions were needed to increase the impact of CDP data while helping users accomplish their objectives. The key to project success was to:

- Identify the solutions that can be addressed;
- Have a well-defined understanding of the information and processes required to provide the solutions; and
- Provide a clear demonstration of the mechanisms by which the data is gathered and information reported.

In the financial arena, XBRL is already required, around the world, by regulators and supervisory agencies. Since 2008, the U.S. Securities and Exchange Commission SEC adopted rules requiring public companies and foreign private issuers to provide financial statements in XBRL, and publish their financial statements on their corporate website using XBRL SEC, 2008. Since then, other regulatory agencies around the world have enacted similar mandates. In Europe, XBRL is now required for external financial reporting by banking regulators and for all publicly traded companies. There is also a relevant gap in the literature on how XBRL can be used for other purposes beyond financial reporting schemes.

In this thesis, we explore the potential use of XBRL to support CDP-related reporting and how XBRL can help in the integration of CDP information into the main control systems i.e. financial information is commonly supported by XBRL in Europe and other areas. This aspect is particularly relevant for a coherent business strategy, as argued by Jansson [34].

Environmental and sustainability reporting initiatives, including the CDP, the GRI and AECA, have initiatives promoting the disclosure and use of their data through XBRL to enhance the adoption and impact of their data for decision-making purposes. It means that XBRL is

becoming in the common denominator between financial and non-financial frameworks; in other words, in the disclosure and use of corporate information.

Regarding the status of XBRL projects, these organisations and their reporting frameworks, it is relevant to know that:

- **CDP** is taking further actions on XBRL, such as updating its annual taxonomies, evolving its reporting system, and working more closely with the software community and their stakeholders on their use of XBRL as a way to increase the adoption of its data.
 - More information: <https://www.cdp.net/en/research/xbrl>
- **GRI** is an independent international organisation that helps businesses, governments and other organisations understand and communicate the impact of business on critical sustainability issues such as climate change, human rights, corruption and many others. GRI's vision is to create a future where sustainability is integral to every organisation's decision-making process. With thousands of reporters in over 90 countries, GRI provides the world's most widely used standards on sustainability reporting and disclosure, enabling businesses, governments, civil society and citizens to make better decisions based on information that matters. Since 2014, GRI has promoted a pilot program to disclose sustainability reporting using GRI XBRL taxonomy. However, as is happening in CDP, they have not yet evolved their systems to be able to accept, validate or publish information in XBRL or another open format.
 - More information:
https://www.globalreporting.org/services/Analysis/XBRL_Reports/Pages/default.aspx
- Since 2006, AECA promotes good practices in sustainable reporting and management to support companies in the disclosure of financial, social, environmental and corporate governance information, making those practices valuable for internal management and decision-making processes. AECA is distinguished by promoting the use of XBRL in non-financial reporting practices and their XBRL taxonomies are acknowledged by XBRL International. Currently, AECA is building the first open data platform to enhance the adoption of XBRL in non-financial reporting, offering advanced analytical capabilities to reporting firms, their stakeholders and general

society. This work includes the collaboration and acknowledgement of several major listed corporations along with key regulatory bodies, such as the Bank of Spain, the Spanish Security and Exchange Commission and the Business Registers.

- More information: <http://is.aeca.es/en>

Despite these advances in non-financial reporting, the reality is that there is still work to do within all of these frameworks, but the provision of XBRL-formatted data concepts allows for the creation of direct linkages between financial and non-financial measures. These technical mechanisms have a real-world impact: they can directly and demonstrably enhance accountability for financial and non-financial reporting alike.

Although different frameworks are using the standard to different degrees at this point, this use is accelerating rapidly, with non-financial reporting experts in agreement that the creation and dissemination of reports in digital form will enhance their comparability and utility. More detail about our contribution on influencing the practices of XBRL within the CDP and its market is presented in Chapter 7.

3.6 Semantic web and linked open data in the corporate world

In recent years, the emerging field of data science in industrial and academic arenas has brought together specialists and researchers from business, statistics and IT disciplines whose work is focussed on a new area: massive and complex datasets in the so-called context of “Big Data”. Although the term has gained popularity in recent years, there is ambiguity regarding its meaning. “High-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insights and decisions making,” is one of the most cited definitions [35]. As a result, many industries are trying to exploit data for competitive advantage [36].

The problem is that the volume, variety and velocity of data have far outstripped the potential of conventional databases and analytical solutions. Traditional data warehousing tools do not support the unstructured data sources and the expectation on processing speeds for analytics. New technologies are required to tackle this new complexity and exploit the new opportunities offered by Data Science.

How to turn corporate data into valuable information to aid in decision-making is one of the questions this research is trying to solve. For that purpose, we agree with Brobst and Rarey [37] on the following fundamental points:

- Access to the information increases the quality of decision-making.
- Developing a superior corporate strategy is a fundamental part of success in a competitive business environment.
- The emergence of a vast amount of information, Big Data technologies and data science techniques influence the execution of a business strategy.

How to exchange corporate data have been proposed by various frameworks to define the different layers in which interoperability takes place, grounded in the following two goals:

- (1) Practical: the ability to exchange and use information between systems. For example, Software as a Service (SaaS) at the application level, syntactic and semantic interoperability.
- (2) Organisational: the ability to coordinate organisations for mutual benefit. For example, through legal and statutory agreements.

In the particular case of the European Commission [38], the European Interoperability Framework is issued to enable effective communication between public administrations in the 28 member states. This framework is considered key in the future European strategic plans towards the consolidation of a unique economy and the Digital Single Market [39]. It covers the practical and organisational scopes described above, distinguishing the following four layers of interoperability:

- (1) Technical interoperability: it involves the definition of technical specifications to ensure the communication between computer systems and services. It includes messaging protocols, data formats, security and services descriptions and properties to ensure the quality of the information, such as consistency, completeness and reliability.
- (2) Semantic interoperability: it covers the technical mechanisms to define the meaning and the relationships between different sources of information, avoiding ambiguous

interpretations (same data, same interpretation). For example, data dictionaries, schemas and taxonomies.

- (3) Organisational interoperability: coordinates the organisational processes to integrate into the internal management and strategies. Involves Memorandum of Understanding (MoU), and the definition of expected level of agreements.
- (4) Legal interoperability: It implies legal agreements to allow the compatibility between different legal conditions in order to access and use information. For example, it involves the definition of appropriate privacy terms to cover the needs of different legal environments involved.

Given the relevance of this framework in the future of the European economy, and for the purpose of this study, we analyse our results in chapter 6 in terms of these four layers.

Thinking more holistically about what could be considered as relevant information from companies beyond their own corporate websites, it is common for Data Science applications to combine data from different sources. That can be the case when considering the interaction between corporate websites and other sources of corporate data, such as social media via Facebook, Twitter and LinkedIn, which enable businesses to interact with their employees, customers, partners, and other stakeholders. Social media maximises the impact of such data by linking or embedding it on third-party sites and allowing collaboration [40]. Along with corporate websites, social media and media sharing sites, it is possible to combine these datasets with those maintained by the public sector, thanks to the so-called Open Data initiatives. Central governments are significant in this respect because as part of their transparency agendas they are freely making a large amount of data available [41]. Well-known examples of these Open Government Data Portals are: data.gov.uk, which is a UK government project launched in 2009 to make non-personal UK government data available as open data, and data.gov in the United States. Open Data is creating both social and economic value by giving away data that was created using public money, and economic stimulus is provided from companies building new products and services around the data [42].

A part of the intention to control the data published on the Web prior to analysing it, a set of best practices for publishing and connecting structured data is in place, so-called Linked Data

[43]. Linked Data connects pieces of related data and information coming from different data sources, structuring information and knowledge. Linked Data seeks to ensure that the information currently available on the internet can be better read and understood by computers, thereby allowing for further exploitation of the data. The adoption of Linked Data is a reality in projects such as DBpedia from Wikipedia, which is focused on converting Wikipedia content into structured knowledge and linking it to other datasets on the Web [44]. Linked Data is also being considered as a key component of Open Data initiatives, particularly as a measure of openness. This checks if the licence is open, if the resource links work and if the resource formats are open and linked to other datasets. A pilot study examining this measure is the openness score developed by the UK Government Open Data site (Figure 9), where each dataset is marked with a score depending on its level of openness. Five stars are awarded as the best openness score to datasets in Linked Data format. Structured data in open formats, such as CSV and XML, are classified as the second-best score, receiving four stars [45]. Social media, Open Data initiatives and Linked Data are important sources for the corporate world to promote better, smarter and real time fact-based decisions. The degree to which data and technologies can become a source of economic value is changing the way business competes and operates. These concepts of semantic technologies are further explored in Chapter 4, and their adoption for effective corporate transparency are developed in Chapter 8.

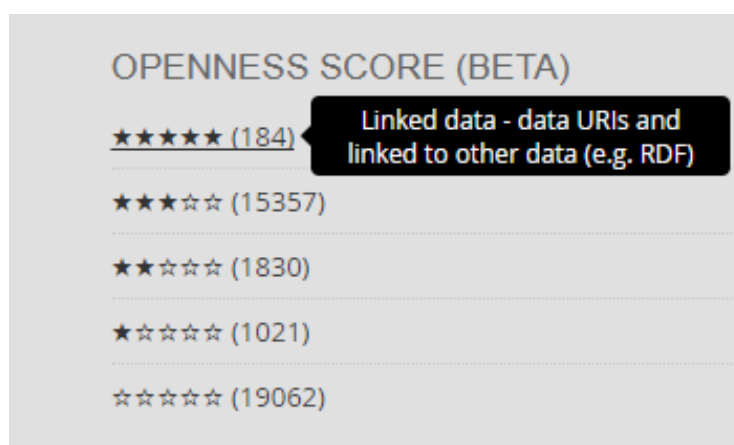


Figure 9. Openness score at the UK open data portal data.gov.uk [Accessed August 2019]

3.7 Research goals

Our work is contextualised in the research areas of accounting information systems and the semantic web. Furthermore, we provide an advance beyond state of the art in these areas:

- Regarding Information Systems and their applications, this thesis supports an innovation in:
 - Reporting standards to overcome environmental-related reporting challenges (Chapter 6).
 - Provenance and accountability of companies, showing how financial and non-financial data from different sources can be linked to holding companies more environmentally accountable (Chapter 6, Chapter 7 and Chapter 8).
 - The use and central development of semantic web technologies to provide transparency (Chapter 8).
 - Methodologies to determine alignment between different corporate reporting frameworks (Chapter 2).
- Regarding the sustainability challenges and their accompanying research issues, we explore:
 - The role that voluntary reporting initiatives can have on the Environmental Information Systems of disclosing firms and consumers of this information (Chapter 4).
 - How to reduce the costs of environmental reporting practices and how to increase the value of environmental data in specific areas of the companies (Chapters 4 and 5).
 - How to better understand the link between environmental and financial implications (Chapters 5 and 6).
 - The role of interoperability (technical, semantic, organisational, legal) to the exchange of environmental and sustainability data/information for better accountability (e.g. policy and regulation) and decision making (e.g. within the business and at policy level) when it comes to sustainability (Chapter 6).

In summary, for the nature of this EngD programme, this thesis is framed as applicable research which is trying to bring answers and solutions to the following goals:

- 1) Market understanding on the use and disclosure of environmental data.
- 2) How to make environmental data more accessible for users?
- 3) How to enable the use of environmental data in combination to financial information from companies?
- 4) How to support the consistency, analysis and data alignment between different reporting frameworks?

All these goals have the objective of improving environmental reporting practices through Information Technologies and so impacting CDP (and wider industry).

Chapter 4. Literature review and related work

The IS research community has not paid detailed attention to the role of IS in the field of environmental sustainability [46]. In summary, these are the gaps identified in the literature:

- The lack of easily accessible corporate data to quickly and accurately inform the management about material issues, which are pertinent during decision-making processes [47] [48].
- The lack of awareness of the impact sustainable business practice and reporting can have on financial outcomes [49][50].
- The inadequate levels of integration of financial and non-financial information within the internal performance, strategy and operational frameworks of an organization [51].
- A dearth of consistency, comparability, reliability and clarity of climate change information emerging from organisations globally [52][53].

Questions arise regarding how IS can facilitate environmental sustainability in business organisations by enabling:

- 1) Better environmental reporting processes: data collection, data transformation, validation and publication.
- 2) An interactive exchange of information between stakeholders.
- 3) An efficient data management process, turning information into decisions: collecting, storing and processing information by appropriate means.

IS research in the field of environmental sustainability remains in its early stages, more focussed on conceptualising and analysing practices instead of giving concrete insights about what challenges and problems exist to move from reporting practices to major actions of environmental sustainability. In general, this work is concentrated on providing those detailed insights in order to be relevant for academics and practitioners who are building IS for the purposes of environmental sustainability.

4.1 About CDP and its theory of change

In 2002, the Carbon Disclosure Project (CDP), an international non-profit organization, was launched as a global initiative to encourage corporations to measure, disclose and manage their environmental risks and reduce their carbon emissions. CDP's theory of change proposes solving some of the world's most pressing problems by enabling transparency on businesses' environmental impact—and the impact of the environment on businesses— through appropriate information flows. Namely, information must go from firms to a broad set of stakeholders, thereby enabling users to make better decisions. In turn, this helps to ensure the long-term sustainability of both the firm and its environment.

As mentioned in Chapter 1, CDP works with influential market players, including 827 institutional investors with assets of US\$100 trillion. CDP encourages companies and cities to disclose their impacts on the environment and natural resources and act to reduce them. For that purpose, CDP runs a global disclosure system that enables companies, cities, states and regions to measure and manage their environmental impacts by means of one of the most comprehensive collections of self-reported environmental data in the world. CDP specifically proposes three objectives to achieve environmental sustainability: (1) reducing greenhouse gas (GHG) emissions; (2) reducing water scarcity; (3) and preventing forest destruction. CDP collects data from large corporations, SMEs and cities through its climate change, water and forest programmes, and offers tools and services to facilitate the use of its data. For each programme, a questionnaire is sent to companies on behalf of institutional investors that endorse CDP's work and support the disclosure requests of each programme.

As a result, CDP now holds the largest global collection of information on primary climate change, water and forest risk commodities, and puts these insights at the heart of strategic business, investment and policy decisions. Over 5,600 organizations, including 81% of the world's largest public companies, now use CDP to disclose vital environmental information to investors and major purchasers [54]. Through their offices and partners in 50 countries, they are considered a major global actor in non-financial reporting.

4.2 CDP in the literature

CDP is considered differential in the data market for growing a unique database with environmental information from companies and cities [55]. In its organisational role, CDP represents a dialogic engagement initiative that facilitates the environmental disclosure by companies and cities with institutional investors [56]. It serves as a corporate governance mechanism for shareholders to influence the firm's environmental disclosures [55][57], as CDP is an influential player that enables a closer collaboration between different stakeholders. Operationally, it is characterised as reducing the asymmetry of environmental information, enabling comparability and a greater level of uniformity of environmental data due to CDP not only assessing companies on environmental disclosure, but also collecting environmental information directly from companies [58].

4.2.1 CDP strategy

The strategy followed by CDP to encourage participation is considered a new stage in the evolution of shareholder activism where investors pressure firms to disclose environmental information [59][60]. The combination of pressure and shareholder activism is the strategy that CDP uses to influence a firm's environmental disclosures [57]. This strategy seems to work according to Wegener [59], who state that CDP's potential success is to encourage companies to disclose environmental information with the expectation of renegotiating emerging social and environmental issues [56]. However, Andrew and Cortese [58] criticised this strategy followed by CDP, accusing it of being a "green capitalism initiative", with a sole concern for investor interests, showing a lack of real benefit to the general public.

4.2.2 The influence of CDP through its reporting initiative

Most authors agree on CDP's success in generating more voluntary environmental disclosure [58][61][62][63]. Wegener *et al.* [64] argue that, for the reporting firms, CDP may represent an inexpensive means to gain positive publicity by enhancing their "green" credentials. However, Lydenberg [65] argues that the absence of mandatory reporting suggests that environmental matters are not a priority for organisations themselves, and therefore investors assume that the information related is not being adequately managed by

organisations. On the other hand, Andrew and Cortese [58] defend CDP as a mechanism of influence for regulatory regimes to try to meet stakeholders' expectations. They explain that, "There is no doubt that the CDP will influence emerging mandatory and self-regulatory regimes because the repository is a source of significant information that can be used by policymakers, educators, academics, investors and creditors". Nonetheless, the authors highlight a lack of reliability with regards to the information managed by CDP, given the fact that the project is not mandatory and companies can omit information or request their responses not be made public. The authors conclude that the lack of policies to make the reporting process to CDP compulsory has a real impact on the consistency and reliability of the information disclosed by the firms.

4.2.3 The influence of CDP data approach to create impact in the market

Looking further into the role of CDP with regard to environmental reporting practices, it is relevant to note that studies identify CDP as an initiative that promotes the standardisation, uniformity and centralisation of environmental data. In addition, it also encourages the harmonisation of methodologies to calculate environmental performance in order to enable the analysis and comparison of this information, informing the companies and their stakeholders' interests and concerns related to these matters [58][66]. Thanks to this integrated approach, CDP is considered an important driver for companies to improve the quality of their environmental reporting practices [58][67][68]. The improvement of data quality has a direct effect on the meaningfulness and usefulness of the information, which has been pointed out as a key influence on decision making in daily operations [69]. The ability to compare, manage, monitor and measure environmental data is recognized by Rankin *et al.* [61] as the main benefit of CDP. According to Andrew and Cortese [58], the benchmarking opportunities provided by CDP data encourage companies to measure and improve their own performance.

Despite these well-known benefits of CDP's data quality and reporting process, other authors find the quality, usefulness and comparability of the data fall short of its desired goals [63]. For example, despite the information centralised by CDP, it is difficult to compare due to lack

of common methodologies to calculate certain factors as well as the lack of precision and the limited level of assurance.

4.3 The role of information systems in environmental sustainability

Companies require solutions to support their corporate reporting practices and obligations. They need systems to respond to social and environmental challenges, reducing their impacts and strengthening their business in the long term, beyond required rules and regulations. Likewise, their stakeholders need better ways to consume and use the business information to make more well-informed decisions. If companies fail to meet those challenges, they are at risk of being driven out of the market [70].

“Business needs long-term approach towards maximizing inter-temporal profits, an active stakeholder management process, and more developed measurement and reporting systems.”[70].

Several research projects regarding Information Systems¹⁰ (IS) investigate how the presentation of information can influence environmental sustainability actions [71][46][72][126]. For instance, what should be the content and form of corporate environmental reporting information, and what types of information are critical to promoting environmental measures and decisions? IS in the field of environmental sustainability is characterised by:

- 1) Providing the necessary information and tools to integrate environmental sustainability aspects into business operations and decision-making processes [127]. For example, proper use of analytical tools and reliable environmental information from companies are important to ensure that key stakeholders such as investors, governments and general society are aware of their environmental impacts and actions. Otherwise, they will be not able to provide the necessary actions and commitments .

¹⁰ An information system (IS) is a combination of people, processes, and technologies that enables the processing of digitised information.

- 2) Ensuring an organisation's reporting practices comply with laws and regulations, and act as a resource for internal management and new market opportunities [128].

In this context, Melville [72] developed the belief–action–outcome (BAO) framework to demonstrate the critical role that IS can play in shaping sustainable practices in organisations, providing the foundation to enable processes and new practices in support of belief formation, action formation, and outcome assessment (Figure 10). The BAO framework is presented as a solution to boost the adoption of IS for environmental sustainability and provide a better understanding of critical issues. Regarding CDP, it plays a role in collecting and exchanging environmental information, but it is lacking with regards to providing practices and IS processes to enable better analysis, monitoring and ease of integration of its information in decision-making systems.

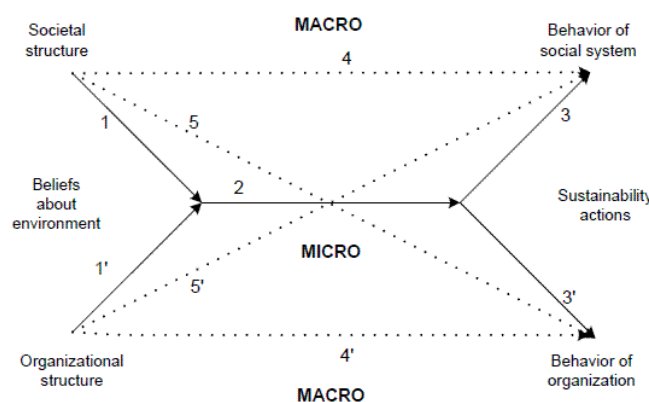


Figure 10. Belief-Action-Outcome (BAO) Framework for IS Research on Sustainability

In addition, within the IS literature are models based on institutional theory that focus on the process by which certain ways of thinking and doing become accepted practice or embedded in institutions [73]. Adela *et al.* [74] in particular found this theory useful for describing how IS could contribute to environmental sustainability in organisations. In fact, the authors developed a conceptual model based on institutional theories to understand how IS can influence human actions in environmental sustainability under different institutional pressures: driven by competitive pressure with other organisations (mimetic), cultural expectations (normative) and governmental laws and regulations (coercive) (Figure 11). With

regards to CDP, they considered these institutional theories at the organisational level but the integration with IS was lacking to create real impact on data-based decisions.

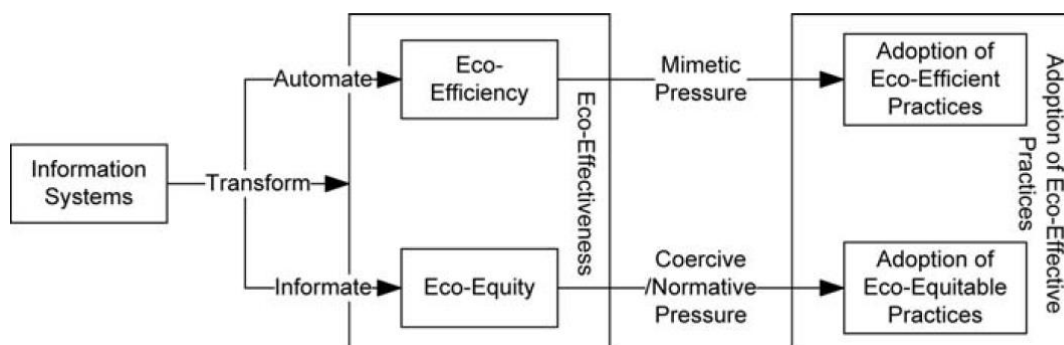


Figure 11. Conceptual model of IS and ecological sustainability

Although these studies call for new ways of examining the roles of IS in organisations, challenges remain regarding the quality, reliability and relevance of the existing environmental data. Thus, reporting practices do not meet the demands of stakeholders because there are several deficiencies regarding the quality, reliability and comparability of this information [75], and there is a lack of adequate IS for engaging stakeholders with this information. In fact, most of the application-orientated studies conclude that current IS adopted by organisations are not designed to capture social and environmental data, and they do not support the data management of all the sustainability dimensions [76].

The following two sections introduce background about corporate disclosure—theories, motivation, opportunities and continuing challenges—as well as the evolution of reporting standards. In this respect, we focus our attention on the case of CDP as a voluntary reporting initiative and as a claimant for better environmental information systems in the context of reporting and management.

4.4 Corporate reporting over the internet

In areas where the Internet is widely available and there is a large number and variety of stakeholders, business reporting has evolved rapidly to include a great amount of content. This now involves not only the corporate website of the firm itself, but also content produced

by information agents such as journalists, analysts, rating agencies, governmental agencies, environmental organisations, scholars, customers, etc. This links with the concept of corporate dialogue [77], understood as a new situation in which the company offers its stakeholders the possibility of analysing the company's public information and activities. These stakeholders are then enabled to express their own opinions publicly in the same virtual space. This ideal situation can be a reality due to enhanced corporate websites and social network sites (SNSs). The corporate dialogue can be understood as a continuation of the traditional corporate disclosure policy, as it implies more dynamic reporting, a combination of mandatory and voluntary information and a more participative way to present and analyse company information (Figure 12).

The modern-day online scenario is characterised by a huge offer of corporate data, but a significant asymmetry exists: it is unlikely an analyst can efficiently process such large amounts of complex data. Reducing such a gap through Data Science is then justified.

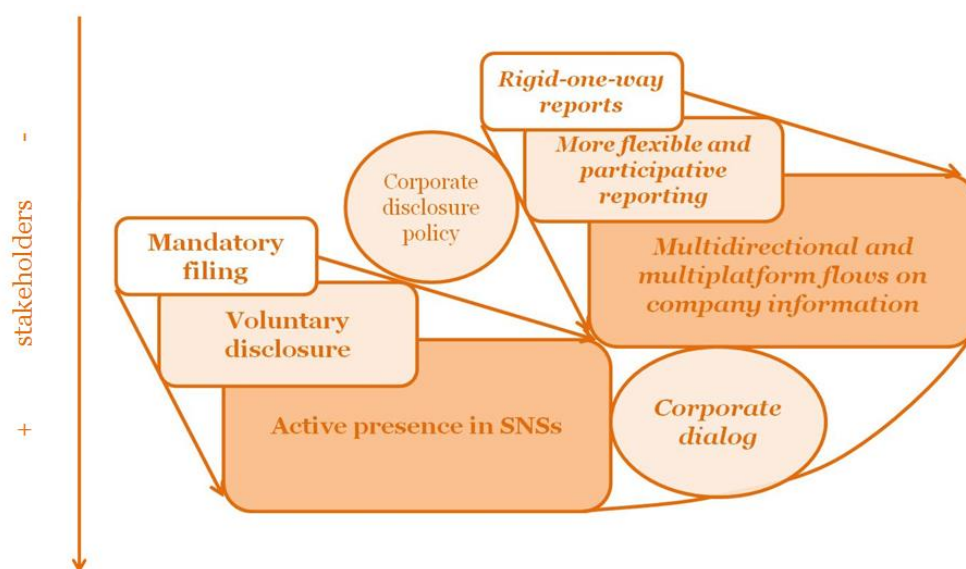


Figure 12. Corporate dialogue as an incremental step from traditional online disclosure

The platform in which this corporate information can be found has also undergone a profound transformation, from periodic paper-based reports to multiple dynamic datasets on the company's website containing HTML, PDF, XML, XBRL, spreadsheets and other formats; official repositories of financial reports; spaces in social network sites; or shared video and

audio content. This data can be relevant to make informed decisions and assess the situation and behaviour of a given company, its results, its image and online reputation.

4.5 Technologies for corporate reporting: XBRL and Linked Data

The current landscape of environmental sustainability data from corporations lacks digital standardisation and connectivity, so we will explore XBRL and Linked data technologies to address both.

4.5.1 Data Standardisation - XBRL

Research on corporate transparency remains relatively scarce. However, academic studies show that there is interest in using information technologies in sustainability reports, at least in part to guide investment decisions [78]. Cohen and Santhakuma [79] recognise that reporting obligations could go beyond having just economic impact.

There is a growing amount of literature regarding the methods and format in which this reporting is disclosed. The pioneering book *One Report* by Eccles and Krzus [80] introduces the idea of using websites for reporting: not just as containers for PDFs, but enabling users to interact with the reports in a much more sophisticated way:

“When the Web is used to provide information, much higher degrees of integration is made feasible. Not being limited to the linear nature of the paper format (...) Every piece of business information (...), tonnes of carbon emissions (...) can have an electronic tag (called metadata) that enables access to this information”. (Eccles and Krauz, 2010:191)

Accordingly, one of these new forms of reporting is based on the use of metadata languages that add electronic tags to every relevant piece of published data, allowing for automated digital treatment. XBRL (eXtensible Business Reporting Language) is mentioned by these authors as one of the most extensively acknowledged standards for that purpose, widely used at the international level to support the reporting of regulated financial information, but just recently considered for non-financial data transmission. Several authors draw attention to the advantages of new forms of communication technologies, in particular for reporting standards like XBRL [81].

The main idea behind XBRL is standardisation. A single taxonomy is created for a specific rule or guideline (i.e. IFRS, US-GAAP, GRI, EMAS, CDP, etc.) and its corresponding statements and reports. The taxonomy is where the rules and data definition is stored, and it is comprised of a set of elements (e.g KPIs) and all the presentation, calculation and logic rules that operate according to that rule or standard. Once created, the XBRL taxonomy is made public as an open remote resource on the Internet. Then, for a specific firm, the proprietary software can create an XBRL instance (the report itself), containing the concrete facts and figures for a certain period. The XBRL instance can be checked against the taxonomy by all parties (e.g. reporting entity, a regulator, or even the public) in order to guarantee its validation. The creation of an XBRL taxonomy implies the agreement of all interested parties. When applied to financial information, the XBRL working groups involve regulators, IT experts, academia and industry; this was also the case for non-financial guidelines like GRI and CDP.

Once the taxonomy is public, the reporting entity must adapt their pre-existing systems to be able to prepare and publish such an XBRL report. There can be multiple final destinations for the report: the corporate website, an official reporting platform or repository, etc. Once in XBRL, business facts are much more accessible to different types of applications for data analysis, allowing users to make quick calculations, ranking and comparisons. The reporting entity itself can also benefit from this digital format for consolidation or internal auditing purposes. Arndt *et al.* [82] explored the use of XBRL for environmental reporting in the context of the Global Reporting Initiative (GRI). The authors defined XBRL as the language which specifies the syntax of a report and can be defined as a number of report concepts, as well as its respective contents. According to the framework approach, reports consist of two levels: XBRL instances and XBRL taxonomies. The environmental report can be understood as an XBRL instance, and the XBRL taxonomy is the set of minimum content and business relationships that the data contained in the environmental report (instance) must respect. Hřebíček *et al.* [83] find the use of XBRL to be essential for GRI reporting in particular, as long as the use of tags is linked to the definition, reporting and transmission of Key Performance Indicators (KPIs) in the economic, social, environmental and corporate governance arenas.

The aim of XBRL is to improve the communication of financial and business information, facilitating the exchange of the information and validation by disclosers and users, thereby

enabling better decisions through an open standard [84]. XBRL is preferred as a standard format by regulators and by companies that use it on a voluntary basis [85]. XBRL is being applied with greater frequency for non-financial data, e.g Global Reporting Initiative G3 guidelines [86] and CDP.

The XBRL Consortium and its several Working Groups are the organisational structure whose mission it is to create each specific XBRL taxonomy; this interdisciplinary group, where experts in the domain and in the technology work to form consensus on which specific concepts should appear in the corresponding standardised report and which business rules apply, help clarification and interpret abstract rules. That was the case for the International Financial Reporting Standards (IFRS) and the United States Generally Accepted Accounting Principles (US-GAAPs), for financial reports, Basel III for solvency reports, GRI G4 and so on.

Despite these developments, the wide application of XBRL was not certain and it experienced some obstacles in different regulatory and industry contexts. Troshani and Rao [87] identify environmental, organisational and technology factors that apply to the challenges of XBRL adoption and dissemination. For example, there is very limited software support for XBRL data consumption, and unfortunately XBRL's proprietary tools lead to vendor lock-in. This is likely motivated because the majority of generated XBRL data is private and non-accessible, and only can be consumed by regulators or supervisory entities. From a technical perspective, XBRL has evolved to support both global regulatory environments and emerging domains such as sustainability reporting. The flexibility needed to do this has resulted in difficulties with regards to standardisation. This is primarily because of the diversity of technical implementations produced by different modelling practices during the taxonomy development phase. In Chapter 8, we explore how our work has contributed to solving some of these challenges to enhance and improve corporate transparency.

Janvrin and Won [88] also studied the issues surrounding XBRL implementation, in particular the extent to which companies are prepared to implement XBRL, and whether software tools and guidance exist to help preparers through the process of creating XBRL-related documents. Through this process, four factors were revealed as worth monitoring to influence the spread of XBRL in business environments: management support and involvement, implementation approach, organisational readiness or expertise, and control

over the XBRL reporting process [88]. Thus technical as well as organisational and managerial topics must be taken into account when implementing XBRL for reporting. The extent to which a certain working environment is appropriate for XBRL reporting procedures is still a key issue. In Chapter 7 we explore these topics with regards to the role that XBRL can play in climate-related reporting.

4.5.2 Data connectivity - Linked Data

In our interconnected world, the information and the technology that enables its generation and transmission are critical. In fact, all companies that compete and collaborate both locally and internationally need an adequate channel of communication with their various stakeholders, including investors, customers, suppliers, employees, governmental agencies, and even the general public. The Internet and the extensive use of corporate websites is the preferred medium to support this communication. The need for interaction between a company and its diverse stakeholders has also led to the adoption of Web 2.0 technologies and social networking sites (SNSs) as a method of communication.

A variety of data can be found on the Web, which is accessible in different formats such as Excel spreadsheets, images, videos, etc. While humans can easily access this data, it is difficult for computers to understand because the information and data presented do not provide further details about its content and context. In recent years the Web has evolved from a global information space of linked documents to one where documents and data are linked. Underpinning this evolution is a set of best practices for publishing and connecting structured data on the Web, known as Linked Data [43]. Technically, Linked Data is the sum of good practices for publishing data on the Web, RDF (*Resource Description Framework*) and vocabularies.

RDF is a standard for describing resources on the web. It permits the representation of relationships between data in triple format; object, subject and predicate, by means of three elements:

- Resources: the things being described.
- Properties: the relationship between things.
- Classes: the buckets used to group the things.

A vocabulary is a collection of terms to describe more properties and classes [89]. There are currently more than 600 vocabularies available¹¹, which can be used to model a Linked Data structure (Figure 13).

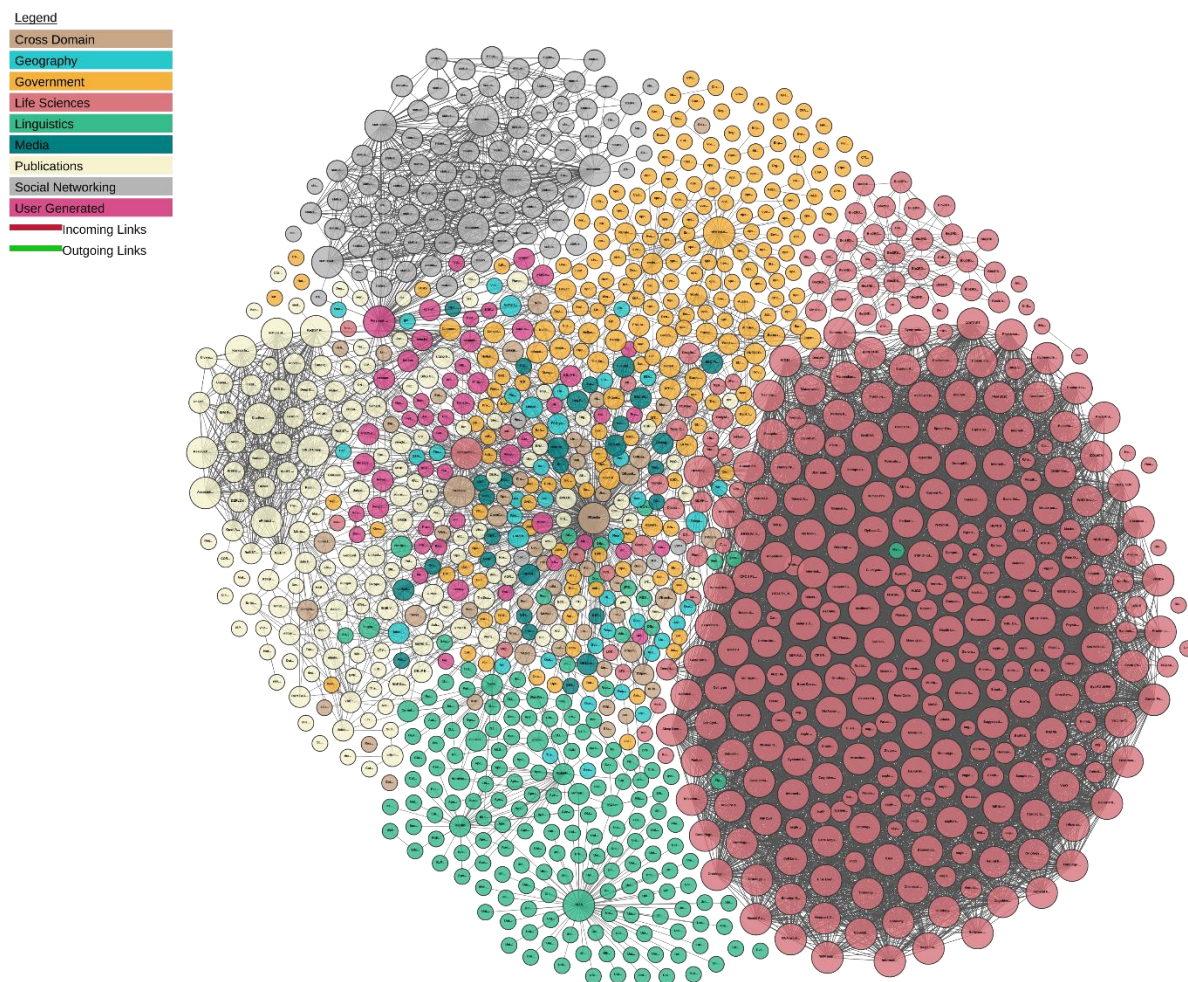


Figure 13. Linking Open Data cloud diagram 2017, by Andrejs Abele, John P. McCrae, Paul Buitelaar, Anja Jentzsch and Richard Cyganiak

The set of vocabularies is centralised by the Linked Open Vocabularies (LOV) initiative and it was adopted as an official project by the Open Knowledge Foundations in 2012 [90]. LOV is an essential resource for any Linked Data developer, as it enables searching the adequate vocabularies depending on semantic needs. The fact that a Linked Data architecture is valid

¹¹ Linked Open Vocabularies: <http://lov.okfn.org/dataset/lov/vocabs> [Accessed August 2019]

to use more than one vocabulary, and there are more than 600, represents an opportunity to express a large range of relationships between terms. The number of vocabularies indexed by LOV is constantly growing thanks to the community's effort.

What can Linked Data offer in the area of corporate reporting? Good practices for publishing information on the Web, RDF and a significant number of vocabularies indexed by LOV. There are several governmental initiatives that use the Linked Data approach for publishing environmental data, such as:

- The German Federal Environment Agency for publishing Environmental Specimen Bank and the Semantic Network Service (SNS) [91].
- The UK Environment Agency¹² for publishing bathing water and natural resources data in the UK.

In Chapter 8 we refer to more research on the use of Linked Data and the semantic web for industrial and public sector reporting.

In our study, vocabularies and Linked Data principles are used as a solid technology to represent and publish corporate data on the web and enable better performance analysis of sustainability, primarily connecting data that have something in common. In this way, more contextual knowledge can be included in sustainable data. These ideas are supported by Cayzer and Preist [92] where, according to the authors, "Sustainability data does not stand alone. Rather, in order to interpret and use this data in a principled manner, one needs a considerable amount of contextual knowledge. Context is the surrounding information that is needed to make sense of a figure."

In Chapter 8, Linked Data is applied as a solution for publishing financial and environmental reports connected with existing company information in the Web, overcoming the limitations in corporate reporting for having this information in a disconnected manner.

4.5.3 Standardisation and Data Connectivity working together

Previous efforts to make XBRL data more interoperable with other data sources and formats have used RDF and OWL ontologies, as well as linking and publishing solutions. The majority

¹² The UK Environment Agency open data portal : <http://environment.data.gov.uk/index.html> [Accessed August 2019]

of these base their examples on transforming financial XBRL taxonomies models into RDF from well-known open government data initiatives, such as XBRL filings available from the SECs EDGAR program. However, none of these studies covers the full XBRL specifications including XBRL 2.1 and Dimensions 1.0. In other words, these studies do not offer a general solution to convert any XBRL report to RDF. For example, Garcia and Gil [93] propose a solution to transform XBRL filings available from the EDGAR program to RDF. Their approach is generic to the XBRL 2.1 specification: simple items, scenarios, segments and tuples data structures. They use US-GAAP reports from 2006 as a case study, which do not use Dimensions 1.0 specification. On the other hand, Kampgen *et al.* [94], propose RDF Data Cube Vocabulary to model XBRL reports as a multidimensional dataset. They exemplified their methods by using 2009 and 2011 US-GAAP reports, whose taxonomy uses XBRL 2.1 and Dimensions 1.0 specification. However, it is unclear how that solution can be generic to other dimensional taxonomies and how the ontology proposed covers tuples, simple items and contextual information modelled with scenarios and segments. There is an experimental initiative, called the Edgar Linked Data wrapper¹³, that provides access to XBRL filings from the SEC as Linked Data. The approach is to publish US-GAAP taxonomies into RDF as vocabularies. In fact, each new US-GAAP taxonomy version means a new semantic vocabulary. This represents a solution to convert US-GAAP reports into Linked Data, but it is not a solution for any other type of XBRL reports, such as CNMV reports. Closer to the sustainability domain, Madlberger *et al.* [95] presented an ontology-based approach using GRI-XBRL taxonomy to build a Corporate Sustainability ontology. However, the result is a content-based approach instead of a metadata conversion, meaning that the solution proposed is not generic to transform any XBRL report to RDF, only GRI reports to RDF.

Authors agree that there are some limitations when representing XBRL data in RDF graphs and as Linked Data due to the lack of formal semantics and inference mechanisms, and there are also difficulties finding correspondences with well-known vocabularies (SKOS, FOAF, etc.). Furthermore, complete architectures for evolving information systems enabling a better financial data integration using Linked Data are proposed in Goto *et al.*, [96] and Kampgen *et al.* [94]. Basically, these solutions integrate XBRL financial data with DBpedia and

¹³ Edgar Ontology Central: <http://edgarwrap.ontologycentral.com/> [Accessed August 2019]

Yahoo!Finance Web API. For the purpose of this study, we also consider their requirements necessary to boost effective corporate transparency:

- To break the barriers which hold XBRL data in isolated data silos of information and vendor lock-in of proprietary XBRL tools;
- To reach a better level of data coverage and data quality; and
- To facilitate a comprehensive picture of company performance.

We distinguish our study from previous work by proposing a solution in chapter 9 to enhance an effective corporate transparency, increasing the adoption of financial and non-financial data and generating impact on decisions. Our central thesis is that in order to create that impact, two components are necessary: (1) Foster interoperability across economic, social and environmental data published in XBRL format and others; and (2) better integration of these combined data in information systems that are part of the decision-making processes of companies, their stakeholders, regulators and supervisory entities. In order to turn corporate data into valuable information for decision making, we focus on the following tasks:

- A generic ontology to transform any XBRL report into Linked Data.
- Interlinking with existing data available in the LOD cloud.
- Data publication via SPARQL (Protocol and RDF Query Language) query endpoint.
- Enabling data contextualization, cross-data-source analyses and data accuracy.

Chapter 5. Stakeholders' study: the needs for achieving environmental sustainability

This chapter analyses the role of IS in the field of environmental sustainability, from disclosure firms to the corresponding effect on key stakeholders. CDP is a particularly relevant case study for these purposes. This chapter is trying to answer our first objective “Market understanding on the use and disclosure of environmental data”.

One of the most interesting and important elements of this chapter is its attention to the different stakeholder groups and the analysis on market behaviours with regards to corporate reporting. Interviews were conducted with 21 CDP members, including companies, investors, governments, academics, NGOs and software/data providers. The results show clear needs to (1) collaborate more actively in environmental initiatives, (2) gain access to a larger and more diverse environmental dataset and (3) have innovative IS solutions to support data quality, data interoperability, accessibility, analysis and visualisation.

5.1 Introduction

Environmental degradation—such as air and water pollution and deforestation—is a serious problem, and one which is often attributed to industrial sources. As a result, companies are under pressure from consumers, investors and governments to take responsibility for their environmental and societal impacts. Therefore it is necessary to make fundamental changes to how business is conducted in order to mitigate and minimise its impact. We refer to this as a business transformation [97]. Our central thesis is that such change can significantly counter environmental degradation, particularly improving the way companies and their stakeholders make decisions considering corporate information. Timely and accurate information can reduce uncertainty and allow decision makers to quickly and continually adjust activities in response to environmental changes as soon as they occur [98].

The details of environmental impacts of modern corporations are increasingly made public through corporate sustainability reports and other venues. The drivers for this include regulatory compliance, reputation enhancement, meeting investor demand for performance

information, and fulfilling a commitment to demonstrate an ethical position to stakeholders. The disclosure of environmental information is mainly influenced by:

- (1) New regulations, which define compulsory corporate information to disclose, including environmental aspects. An example is the European Directive 2014/95/EU for non-financial reporting, which in 2017 applies to public interest organizations with more than 500 employees, representing approximately 6000 large companies in Europe. The Directive defines a set of information to be disclosed and suggests a set of reporting frameworks to help to comply with it, such as the Eco-Management and Audit Scheme (EMAS) and International Organisation for Standardisation's (ISO 26000). How to apply the Directive is a responsibility at national level: the Directive has to be adapted by each EU member into local legislation and no later than December 2016.
- (2) Global agreements on climate change, with initiatives like the Paris Agreement that seeks to keep the increase in the global average temperature below 2°C, as a result of the United Nations Framework Convention on Climate Change [99]. Nearly 200 nations signed it, and it is the responsibility of each government to decide which measures to impose to reach such targets.
- (3) Voluntary reporting initiatives, which help companies to support their reporting practices by identifying relevant information to disclose and manage. An example of voluntary frameworks for sustainability reporting is the Global Reporting Initiative (GRI) and Dow Jones Sustainability Index (DJSI), and CDP, which only focuses on environmental concerns. The latter is the primary focus of this study.

The objective of this chapter is to identify the role of IS in environmental sustainability and the socio-technical properties that can have a significant influence on the IS of a diverse range of stakeholders, from the disclosure of information to the subsequent management. CDP has been selected because it runs a global disclosure system that enables companies, cities, states and regions to measure and manage their environmental impacts by means of one of the most comprehensive collection of self-reported environmental data in the world. Thus, it can potentially influence an organisation's business decision-making processes and daily operations. In other words, boost the business transformation by improving the way in which decisions are made.

5.2 Background

This work mostly relies on Elliot's framework [97] for IT-enabled business transformation to evaluate the role of IS initiatives in mitigating business-related environmental impacts. Particularly, how IS could increase process efficiencies and information effectiveness towards better decision making.

Elliot [97] presents a socio-technical system where the IS are the central solution to improve the environment by supporting the exchange of information and knowledge between firms and their stakeholders (Figure 14). By doing this, IS are allowed to change human behaviour by taking more actions and decisions considering environmental impacts. To reduce environmental impacts, it is not only necessary that companies take certain actions, but also that firms and their stakeholders (society, government, industry and alliances, organizations and individual groups) adopt an integrated approach sharing understanding and activities, IS innovation and research. This integrated idea is shared by CDP's theory of change through the engagement of influential market players and the incentivisation of better environmental decisions and actions through its focus on transparency and disclosure.

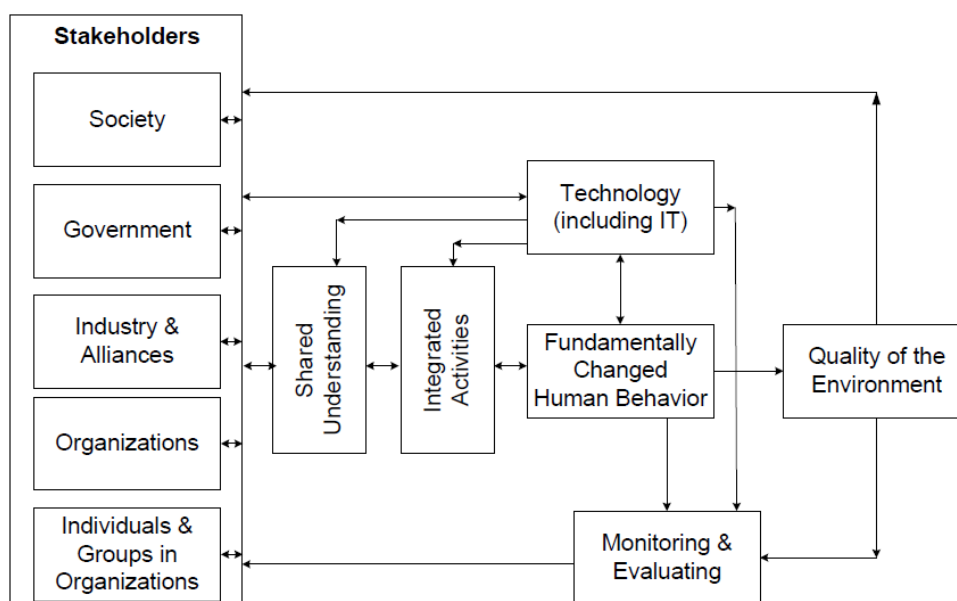


Figure 14. A socio-technical system to impact on environmental sustainability by Elliot [97]

Pressure from stakeholders influence companies to adopt more sustainable practices and inform a company's decisions and actions. Thus, companies need to develop Information Systems able to monitor, analyse, collect and exchange environmental information. This model identifies five groups of stakeholders as the principal contributors to environmental impacts and IS are placed as the central mechanism to support communication among stakeholders, allowing them to share their understanding of environmental challenges in order to change human behaviour to take more actions and decisions considering environmental impacts. IS are also considered as the instruments to monitor and evaluate that behaviour, enhancing the awareness of environmental impacts.

Elliot's framework is developed along with a set of holistic hypotheses. Seeking to explore the relationships between humans, human behaviour, the environment, and technology, the author highlights that it is important to examine the different mediating and moderating effects of technology (including IS) on the natural environment.

After the literature review presented in Chapter 4, our understanding is that IS research tends to limit its focus to business organisations without fully considering the perspective, needs and interactions of other stakeholder groups. This chapter concentrates on providing a broader stakeholder perspective.

5.2.1 CDP as an environmental reporting system

In order to understand and contextualise some of the results of the study and discussions, certain processes of CDP are clarified below, notably the climate change questionnaire¹⁴, disclosure and scoring processes.

The climate change questionnaire requests quantitative and qualitative information about a company's:

- (1) Carbon emissions, to measure the impact of the company on the environment;
- (2) Strategy and governance structures;

¹⁴ Climate Change questionnaire: <https://www.cdp.net/en/guidance> [Accessed August 2019]

- (3) Risks and opportunities, as a reflection of the level of awareness to climate issues; and
- (4) Management actions to mitigate risks, negative impacts and potential opportunities.

Every year CDP uses public consultations through online questionnaires and industrial focus groups sessions to identify the information considered as critical to making changes in business operations, environmental policies and society behaviour in order update the questionnaires with more relevant information.

Beyond managing the content of the questionnaires, collecting and centralising the data, CDP is also responsible for scoring each response according to their scoring methodologies¹⁵. These are publicly available to ensure complete transparency. The scores measure two things: (1) the level of disclosure, and (2) the performance of each organisation that has responded. The disclosure score represents the level of transparency, which is the result of evaluating the number of questions answered and the depth and breadth of each answer. This score is normalised to a 100-point scale, and when the disclosure score is greater than 50, the performance score is calculated. The performance score measures the level of awareness, implications and quality of each response, represented by a letter from A to E.

Through its scoring methodology, questionnaire management, tools and data centralisation, CDP is recognised as one of the leaders in the environmental reporting landscape, offering expertise, transparency, information and mechanisms to generate environmental awareness in business and investments.

5.3 Research questions

Given the main goal of this thesis, we consider IS as the foundation on which to build potential solutions to reduce environmental degradation by facilitating better internal management, decision-making processes and more actions from:

- Local and global governments, who introduce new policies and measures to incentivise the reduction of negative environmental impacts on cities and businesses. For instance,

¹⁵ CDP Scoring Methodology: <https://goo.gl/gtPZpe> [Accessed August 2019]

making new policies to control energy consumption and pollution emissions by allowing access to improved data.

- Investors, who consider environmental data as part of their investments analysis.
- Businesses, influencing their strategies and decisions. For example, investing in suppliers with better environmental strategies or reducing the energy consumption in offices and/or product development processes by having real-time control of the energy consumption and water waste through sensors [100].
- The general public, in their day-to-day decisions. For instance, offering information about the level of environmental impact in dairy products, controlling energy consumption and water waste at home using sensors, providing real-time information about air pollution in cities, etc.

In alignment with the multidisciplinary model proposed by Elliot [97], we are seeking to investigate the use and effectiveness of environmental information among different markets and societies. By gaining a clearer understanding of current drivers and applications of environmental information and practices, and framing the complexity of decision-making processes among stakeholders. In particular, the evaluation of certain stakeholders and technology hypotheses proposed by the author were included in this study.

The research questions take into consideration the role that CDP offers as a relevant case study to explore environmental information systems within a network of actors—disclosers, providers of systems, users of data and other interested parties—in the context of reporting and management.

Considering the main goal of this study, the place of IS according to Elliot's framework [97] and the role of CDP, we propose the following four research questions for this qualitative study:

RQ1. What are the drivers of the Environmental reporting practices?

This research question seeks to generate two different lists from the participants. One is explanatory factors that motivate stakeholders to disclose and use environmental data, and a second will detail the dependency factors, identifying how CDP can become more relevant to decision-making processes. This enables us to identify why companies disclose their environmental performance to CDP, and why companies and their stakeholders use that data.

RQ2. What is the level of detail of environmental information required for stakeholder's decisions?

To address the second research question, we frame the complexity of decision-making processes, asking to identify three aspects: the volume of data, frequencies and data format consulted by the experts for their relevant decision-making process. This allows the relevant properties of information that the participants are using to be recognised, and the dependencies with other sources of information. The classification of complexity is built on the idea of the 3Vs (volume, velocity and variety), coined by Gartner [101], which defines the attributes of a Big Data problem. That approach facilitates the synergy with innovative IS. In addition, it tries to solve one of the key issues of uncertainty highlighted by Elliot [97]: *What is meant by environmental sustainability?*

RQ3. How to make environmental data more accessible to users?

Unlike the previous two research questions, the third question asks experts about their recommendations regarding data practices to extend the utility of environmental data among stakeholders. It affords us the opportunity to understand how companies and their stakeholders are using environmental data and other datasets, their frustrations with the status quo and ways to improve current practices. Also, aligned with key issues of uncertainty identified by Elliot [97]: *What are the major challenges?*

RQ4. How IS could support environmental reporting challenges and opportunities?

We address the fourth research question in order to generate a list of IS requirements and those desired to enhance the use and utility of CDP. We should be able to address it looking

at the results of RQ1, RQ2 and RQ3. It is aligned with the last two key issues of uncertainty addressed by Elliot [97]: *What is being done about these challenges?* and *What needs to be done?*

These research questions were selected as the framework to structure the interviews because of the focus on obtaining a better understanding of where the value of environmental aspects lies. The idea is to identify practices, potential challenges and barriers to confront to change behaviours and achieve environmental impact.

A list of interview questions can be found in Table 2.

List of interview questions

Section 1. Considering CDP data - Why?

1. Broadly, for what purpose do you use CDP data?
2. What analysis or decision making does your use of CDP data support?

Section 2. Understanding decision-making needs - What?

3. What CDP data do you use, and how often do you use it?
4. What other data do you use alongside this to support your analysis/decision making?
 - How often do you use each of them?
 - For what type of decisions are each dataset relevant?
5. Are you using any of them to cross-correlate with CDP data? What do you intend to achieve with the results of that analysis?

Section 3. Extend the effectiveness of CDP data among stakeholders – How?

6. How do you analyse the data currently?
 - What tools do you use in this process?
 - What form do your results/outputs take?
7. What would make your work easier or more effective?
 - What barriers/cost/frustrations/time wasters do you experience ?
 - What data would you like to have that you do not?
 - Can you imagine tools and functionality that would help?

Section 4. The future

8. Do you have in mind any applications, practices or features that CDP could feasibly introduce in a period of 2 to 5 years to maximise the consideration of environmental information in the decision-making process?
9. Can you imagine other ways that CDP data could be used by you and your organisation?
10. In an ideal world, what more would you like to be able to do?
11. Do you want to add any other information?

Table 2. List of interview questions

5.4 Results

The results of the interview are presented and discussed within the following sections. Based on the results, we identified a list of relevant factors that impact on the adoption of environmental information for decision making. These are summarised in table 3.

5.4.1 Drivers

In this section, we present the factors identified that encourage the disclosure of environmental information to CDP and why its data is used.

5.4.1.1 Drivers for companies as disclosers and users

As disclosers of information, companies consider CDP relevant because it offers an environmental reporting mechanism connected with the interests of a network of actors. *"We are responding to CDP, as it is closer to the interest of a broad range of our stakeholders."* [Company 1].

	DISCLOSERS		USERS			
	Drivers	Constraints	Drivers	Constraints	Improvements	Impact
COMPANIES	<ol style="list-style-type: none"> 1. Be part of a network of actors 2. Be comparable 3. Strengthen communication with stakeholders 4. Learning space for leaders and competitors, to get best practices 5. Reputation: be top company in environmental domain 6. Help in the disclosure of environmental information 7. Long-term strategies 	<ol style="list-style-type: none"> 1. Expertise and time required 2. Incoherences in the questionnaire 3. Lack of engagement across organisation 3. Data required- qualitative versus quantitative 4. Difficult to disclose certain information due to privacy terms 5. The annual updates of CDP questionnaires and scoring methodologies 	<ol style="list-style-type: none"> 1. Introduce environmental strategy 2. Get best practices from companies with better scores 3. Competitor analysis and understand their position 4. Improve their CDP responses to get better scores 	<ol style="list-style-type: none"> 1. Time consuming 2. Data quality 3. Cost of CDP tools considering use 4. Difficult to engage other areas of the company to consider CDP information and best practices in environmental matters. 5. The annual updates of CDP questionnaire 	<ol style="list-style-type: none"> 1. Better raw data: clean, consistent, interoperable with external systems and better formats easy to consume by machines and users. 2. Tools to allow comparison of common CDP datapoints across other reference frameworks such as DJSI, GRI. 3. Tools and resources that help to deal with CDP data. 4. Online help in disclosure submission tool. 5. Annual disclosures should be provided in a format ready for publication. 6. Introduce more social issues in questionnaires. 7. Facilitate the link with other similar data initiatives where CDP shares are reporting aspects. 	<ol style="list-style-type: none"> 1. Introduce environmental strategy 2. Enhance company's reputation 3. Improving environmental understanding

	USERS			
	Drivers	Constraints	Improvements	Impact
INVESTORS	<ol style="list-style-type: none"> 1. To evaluate companies regarding transparency and risks 2. To create ESG methodologies in order to facilitate companies evaluation 	<ol style="list-style-type: none"> 1. The CDP data consumption and analysis, due to: <ol style="list-style-type: none"> 1.1 Data format and data structure 1.2 Lack of solutions to integrate with internal systems 1.3 Responses in different languages 2. To do qualitative analysis 3. Engage more companies to do better disclosure to CDP 	<ol style="list-style-type: none"> 1. Better data format 2. Introduce data verification process 2. Tools which allow manipulation and data download (i.e. queryable tools). 3. Facilitate the integration of CDP data with internal systems to allow better integration with internal processes. 4. More integration across CDP datasets (water, forest and climate change). 5. Facilitate the link with other similar data initiatives where CDP shares are reporting aspects. 	<ol style="list-style-type: none"> 1. Companies evaluation 2. Assessing companies for potential inclusion in mutual fund/ investments recommendations
ACADEMICS	<ol style="list-style-type: none"> 1. For teaching 2. Research on: <ol style="list-style-type: none"> 2.1 Regulatory risks and policy. 2.2 Environmental accountability 2.3 Financial risks related to environmental aspects 2.4 Social network effects in CDP performance 2.5 How the type of questions drive to specific answers, e.g. CDP questionnaire, pathway 	<ol style="list-style-type: none"> 1. Lack of comparability across companies and countries due to the different methodologies 2. Data analysis due to data incompleteness, updates over years, samples are incomplete 3. Data quality: data is not clean, and there are inconsistencies 4. The cost to get CDP data considering use 5. The lack of comparability and completeness on certain questions 6. The way CDP makes its data accessible and consumable 	<ol style="list-style-type: none"> 1. Facilitate the content analysis via pre-packed tools. Searchable database: Doing your query allows final results instead of comparing multiple Excel spreadsheets. 2. Introduce data verification process. 3. Make the disclosure of certain information compulsory to allow proper analysis, and define methodologies in a more standard way. 4. CDP needs to take a decision about their position in the market. Whether CDP needs to collect data, or whether it wants others to do that and experiment with innovative software devices and toolkits. 5. The CDP challenge is about working to get high-quality disclosure but then working to think about how users and others stakeholders relationships work to improve their presentation about the material and for different users groups. 6. Make CDP data more comparable. 	<ol style="list-style-type: none"> 1. Influence in research and teaching

		<p>7. Data analysis is complicated due to: CDP data quality is an issue. The sample is not complete. Data with links not available. Updates in questionnaire over the years.</p>	<p>7. Data in a sort of format where the text is easily extractable and easy to analyse. 8. More integration across CDP datasets (water, forest and climate change). 9. CDP should include more data checks, facilitating the link with other similar data initiatives where CDP shares reporting aspects. 10. Make sure the data is properly collected and clean. 11. Promote academics publications and show how they are used. 12. Work more closely with universities on projects. 13. Build a community with academics doing work through CDP with CDP data.</p>	
GOVERNMENTS	<p>1. Get visibility about companies' and cities' actions 2. Identify environmental targets and monitor companies against their commitments 3. Credibility, recognising the expertise and knowledge of CDP</p>	<p>1. Data in more frequent basis: annually, quarterly 2. Data accessibility 3. Data quality</p>	<p>1. Facilitate the contextualisation of the data, providing more representative company and city data. 2. Better data formats. 3. Align strategies with other reporting frameworks such as GRI and SASB to facilitate cross-checks. 4. Offer an open platform that allows companies and cities to update their commitments. 5. Get access to CDP in real time 6. Collect data from investors to get a good understanding of companies, investors and cities about what amount of money is invested to climate change actions. 7. More sector specific modules and scores.</p>	<p>1. Influence environmental policies at national and international level. 2. Assess national companies on environmental strategies.</p>

NGOs	<ol style="list-style-type: none"> 1. To raise the awareness/influence of environmental stewardship with human impact on business decisions 2. CDP data is deep and detailed 3. To strengthen communication with the public 	<ol style="list-style-type: none"> 1. Data analysis, due to how some questions have been asked 2. Not able to understand the impact of environmental issues beyond companies' decisions 	<ol style="list-style-type: none"> 1. Larger samples. 2. More user-friendly databases to encourage companies to do more. 3. More tools to display the information visually and for analysis. 4. Include more links between CDP data and social impact. 5. Include data verification processes to ensure the reliability of CDP data. 6. Summarise the CDP data in a dashboard to facilitate the understanding of the relevant data. 	<ol style="list-style-type: none"> 1. Raise awareness on environmental issues with human impact on private sectors
DATA/SOFTWARE PROVIDERS	<ol style="list-style-type: none"> 1. To include more value in software products for companies and investors. 2. To improve services on benchmarking and disclosure tools 	<ol style="list-style-type: none"> 1. Difficulties to match CDP with other datasets 2. Keep offering continuous analysis tools using CDP data due to CDP questionnaires updates. 3. CDP data format 	<ol style="list-style-type: none"> 1. Have CDP data in database format 2. Reduce the number of open questions and facilitate the comparison of the data. 3. More integration across CDP datasets (water, forest and climate change). 	<ol style="list-style-type: none"> 1. New opportunities in corporate sustainability and investors products

Table 3. Classification of the results per interview coding

Companies agree that CDP enables them to be compared to their peers and to strengthen communications with their stakeholders. *"We have ended up with the conclusion that we want to show our stakeholders what we are doing, and somehow we had to find the right channel to communicate with them. In the end, we considered CDP as the preferred channel of environmental reporting, because it is advertising itself. CDP has a credential, and that is a big selling point"* [Company 3].

Apart from that, companies also consider CDP as a learning space; they use CDP data disclosed by other companies to learn about the best practices and strategies of market leaders and competitors. *"We are looking for good ideas and practices that can be useful for us and if the rest of the companies are doing that, we should do it as well"* [Company 4]. *"If some responses help, and bring us some ideas to move forward, we use them. That has been the philosophy of CDP, do things transparent in a way that each one can learn from others"* [Company 5].

Disclosure to CDP is also a reputational matter for companies: *"We disclose to other reporting initiatives including CDP to be more sustainable and be the top company in the environmental domain"* [Company 1].

From the perspective of data users, notably companies who use CDP data as a guide to complete their own annual disclosures to CDP, they try to find the best answers to specific questions from leaders to get the best score, driving them to take actions on environmental strategy. *"We analyse each question, looking at best practices in companies with better scores. We try to identify their strengths and evaluate what the impact can be on our strategy if we decide to introduce some of them. We get very nice ideas that can be adapted to our company"* [Company 2]. At the same time, we found that some companies only complete the questionnaire to generate a CDP score, mostly to achieve reputational impact and comply with investor pressures, without any particular follow-up actions.

Most of the companies tend to use CDP data to conduct a benchmark analysis against their peers and competitors, and also to understand how competitors position themselves in the environmental domain.

5.4.1.2 Drivers for investors

Investors use CDP data as part of their evaluation criteria to evaluate companies regarding transparency and risks, analysing how environmental actions are integrated within internal management. *"We have historically used CDP data to gain a better understanding of how climate change is a risk for companies. We are interested also in understanding how others are utilising climate change as a business opportunity"* [Investor 2]. *"We always use CDP data to get a deeper understanding on mitigation of climate change."* [Investor 3].

Investors use voting methodologies, which is an approach to investing by taking voting actions on companies where ESG engagements, risks, financial properties, opportunities and impacts have been unsatisfactory. These methodologies are used to form their investment criteria and judge companies. CDP information is also beginning to influence potential investors' decision-making processes. *"CDP data is part of the information involved to judge companies as part of our voting methodology. We specifically expect that companies disclose information on carbon emissions. So we support particular votes in disclosure."* [Investor 1].

5.4.1.3 Drivers for government

Governments are interested in understanding the main drivers of environmental degradation at the national level in order to assess national companies on environmental strategies and develop environmental policies. They found in CDP a unique data source that can provide visibility on the actions companies and cities are taking, consolidating the information into one, internationally accessible location. They primarily use CDP data to identify environmental targets and track a company's progress against their commitments. *"I am looking at the goals and targets of national private companies regarding GHG emission reductions; I am looking at how they achieve their targets. I am looking if these targets are very ambitious regarding our universal goal of remaining below to the 2 degrees of global warming....Even though it is difficult regarding these targets of remaining below 2 degrees before 2020, I am trying to judge if the private sector targets are ambitious enough to achieve this target"* [Government 1].

For them, CDP data is becoming the principal resource to evaluate how realistic the private sector targets are regarding GHG emissions reduction, as certain information needed cannot be found in other corporate data sources. *“With the CDP data, I can analyse and see the progress of individual companies year by year, which is harder to find in the corporate reports. Because in the corporate reports, the companies only put information which is valuable to them, and they do not highlight their progress in GHG emissions reduction, so we cannot find that relevant information in the corporate reports.... That is why CDP reports are very so important for us”* [Government 2].

Governments trust in CDP. They recognise the different knowledge and expertise that CDP has acquired over the years from working closely with companies and cities across the world. In addition, CDP has earned its credibility by being one of the data providers to platforms like NAZCA¹⁶, which is a reliable source for national and international governments.

5.4.1.4 Drivers for academics

In contrast, we found that academics use CDP information mostly for teaching in business fields and to carry out research in diverse areas, such as environmental regulatory risks and policy confidence, environmental accountability, the impact of environmental aspects on financial performance, social network effects and disclosure techniques (for instance, the way in which a question is asked can influence the answer). The general perception according to academics is that CDP has become more important for academia over the last five years, looking at the rate of growth in scientific publications with reference to CDP. *“In the accounting literature, over the past 5 years a greater number of papers are starting to write about CDP data and researchers are starting to focus on climate change more. So that is why I started to be interested in CDP and its data”* [Academic 4].

5.4.1.5 Drivers for NGOs

NGOs are more willing to use CDP data to raise awareness of environmental stewardship with human impact on business decisions. They recognize the value of CDP information because it is

¹⁶ Nazca portal: <http://climateaction.unfccc.int/> [Accessed August 2019]

deep and focuses on environmental issues with a high level of detail. In fact, they admit that there is no other data initiative providing such information on the environmental initiatives of SMEs and listed companies from across the world. Given the reputation of CDP and the possibility for NGOs to participate closely with CDP in the annual updates of the questionnaires via public consultation, many use the information to strengthen their public communications. *"We look at CDP data to analyse what actions companies are taking on community engagement and also to identify the links between human rights and water and sanitation. So we want to use this data to try to get a baseline of what companies are already doing. We also wanted to use it to try to strengthen the business side and strengthen our messaging around why companies should care. [NGO 1]"*

5.4.1.6 Drivers for software/data providers

For software/data providers, CDP data is being used to add value to their financial and corporate sustainability products offered to companies and investors. These software products assist companies in corporate reporting and footprint lifecycles and financial risks. Benchmarking and reporting tools are two common software products developed that use CDP data. The main focus of benchmarking tools is to enable companies to be compared against competitors and peers under a set of financial, social, governance and environmental aspects. *"In the benchmarking tools, we have different sources of data, but the CDP is the best at the moment for us. We have some data for benchmarking analysis, like European pollution registry, where some companies report data on CO2 emissions and other gases"*[Software Provider 2]. *"Our clients use our tools to compare themselves against their peers and competitors, so they can do averages per sector, and compare electricity consumption per employee or staff from one sector to another"* [Software Provider 1]. In the case of reporting tools, they try to facilitate the disclosure of voluntary and mandatory corporate information, and CDP is offered as a voluntary way to disclose information.

5.4.2 Constraints

We use the term "constraint" in this study to refer to a barrier or difficulty that the interviewees identified in order to disclose and use the information offered by CDP. In terms of disclosure,

companies found it difficult and time-consuming to complete the CDP questionnaires, due to detailed information being requested that requires very specific knowledge of environmental matters. The type of expertise required is not usually found in the departments that typically deal with corporate disclosure. *"The CDP questionnaire is the most time consuming to complete. Although we do have now a proper process to complete the questionnaire, there are a lot of specific breakdowns that ask for detailed information about carbon emission, energy use and other information that are challenging to get. Another problem is how CDP requests things in units, which is different in respect to other initiatives that we report. We need to put additional effort and time to translate our data into how CDP requests it"* [Company 2].

The way in which some questions are asked is criticised by some companies. For example, some of them found questions ask for the same information more than once. Regarding the level of detail required for each question, we identified disagreements between companies. Some of them prefer closed questions rather than open, arguing that the open questions decrease the quality and comparability of the data. *"Comparability is one of the features that we consider more important to make accurate decisions, and CDP should consider introducing a major number of data points ready for analysis, instead a large amount of qualitative information that makes comparability difficult "* [Company 2]. On the other hand, other companies demand that more questions should be open, complaining that there is too much emphasis on numbers and that they are not able to capture all the data to show the value of companies in the environmental domain. *"It is not always possible to tell a story with the data, and if you are looking only at numbers, there are not enough to tell everybody what you actually want to tell"* [Company 3]. In addition to that, companies feel the need to engage other areas of the business during the disclosure process. However, they found the structure of the required information made it difficult to get collaboration from other departments, such as Human Resources, Customer Services or Finance departments.

As users of CDP data, all the participants agree that CDP data is difficult to analyse and use for the following reasons:

- **Data format:** CDP offers information in Excel, PDF and Word, depending on membership. The main criticism is that none of the formats and data structures are easy to analyse, consume or integrate with internal systems to facilitate their study. It always requires additional effort to translate the data into something easier to deal with. *"We recognise that the value of CDP data is unique. However the data comes in a format easy to read but difficult for proper analysis"* [Investor 1].
- **Data quality and comparability:** Incompleteness, inconsistencies and the availability of the data on an annual basis are the main data quality issues highlighted by most of the academics and some governments, investors and companies.
 - a. Incompleteness in terms of information not disclosed by companies, which is essential to compare and interpret certain values. *"CDP is doing a good job on companies' participation and standardising questionnaires, but regarding accountability is weak, many companies leave sections blank, which affects the data quality, and even if they are punished in the scores, they are still happy to do that "* [Investor 2].
 - b. Inconsistencies, in terms of wrong quantities and external links that are referenced but not available, which is part of the information disclosed, are also reported as issues. *"At the moment that you are in the exercise of analysing CDP data in depth, you start finding out that the links provided do not exist and certain numbers reported lack of units and precision. These issues make CDP data an unreliable source of information"* [Academic 1].
- **Lack of proper tools and cost:** CDP offers some analytical tools to facilitate the benchmarking analysis of its data. Companies and investors recognise that although they do not need the tools on a daily basis, they would be of benefit during the disclosure period and when the CDP reports and scores are published. The majority of the companies and investors demand better performance and functionalities. *"CDP offers too much data which is difficult to digest. More resources are needed to go into detail into the data and get all value"* [Company 1]. Some companies and academics found these tools very basic in terms of analytical features. Most of the companies found the cost of the tool too high considering their usefulness. *"The*

cost of the tool is the same for every company. The cost for a big international is the same for the small companies and it is quite expensive for us. It should be different prices per company, depending on the size of the company. We only use the tool once, when I am completing the questionnaire, which itself takes a lot of time and effort" [Company 3].

- **Data accessibility:** Academics, governments, investors and companies found the lack of a tool that allows them to access, filter and download data themselves to be a barrier.
- **The use of local language:** Investors and academics mentioned that having responses in local languages makes the analysis of the data difficult, if not impossible.
- **The annual updates of the questionnaires:** CDP updates its questionnaires annually in order to improve the quality and relevance of its data. However, all the interviewees found this a potential barrier to doing continuous analysis, due to a lack of information that clarifies what the differences are across the years. *"Taking data from year to year is problematic; CDP asks different questions every year, and we do not know how to find certain patterns, so we need to do an analysis on previous years to understand what kind of differences are in the questionnaires"* [Company 2].
- **Incomplete samples:** Investors and academics claim not to have the complete stock market participation, which is relevant for their analysis and research.

5.4.3 Improvements

The analysis of the results suggests five types of improvement to increase the adoption of CDP: better raw data, more integration, tools, questionnaire improvements and actions to strengthen stakeholder relationships.

Regarding the raw data, all the participants asked for improvements on data quality issues in order to make sure that the data is properly collected and clean. In general, they suggest that CDP must introduce data checks to avoid inconsistencies during the submission process and should make the disclosure of certain information compulsory to allow a proper analysis. For instance, CDP should make the disclosure of the emission methodology used to calculate the carbon emissions reported compulsory. Currently, some companies fail to complete the

methodology and just include the final carbon emissions number. Such a step would allow the accurate comparison of data over the years and between companies. Data verification is also mentioned by investors and academics as a way to enhance the reliability of the data: *"If there is another source of data where CDP information appears, then CDP should introduce some verification in order to enhance the reliability of its data"* [Investor 3]. Another issue in need of improvement is allowing the consumption of the information by users and their internal systems. This has been identified as a constraint for proper analysis of the information. *"What we would be interested in is to be able to upload the data into our internal systems. As ESG [Environmental, Social and Governance] information is becoming a more and more important part of our business, it requires more integration in our processes, and more automated use of our data"* [Company 2].

CDP needs to offer more tools to facilitate the access and analysis of its information. However, it looks like the primary interest is to have tools to filter and download CDP data in specific formats—most of them mentioned Excel—that would allow them to import CDP data and analyse it using their own tools. *"What would be helpful for me is to extract the data in Excel and then manipulate by myself. In that way, I could compare multiple companies, extracting questions for a certain number of companies"* [Company 2]. Only a few companies mentioned improvements to the current analytical tools offered by CDP in order to enhance visual capabilities and analytical techniques. However, most companies recommended that CDP to offer their tools with different prices depending on company size or to make them available for free. Governments indicated an interest in improving the disclosure tool, suggesting the possibility of allowing companies and cities to update their commitments regarding targets and progress on a more regular basis (e.g. monthly or quarterly). In addition, companies demanded online help with the questions to improve their understanding of the data requested. A few academics suggested that CDP should offer pre-packaged tools to facilitate the analysis of its text content.

In terms of data collection, governments considered investors as the most relevant stakeholders to request information through CDP as, through their pressure, information about how much money is invested in climate change actions can be obtained. In addition, they would like to see more sector-specific data and rankings from companies in order to get more realistic figures on

GHG emission reduction. *"It is difficult to evaluate how ambitious the targets are, because the ambition depends on the technical barriers to each sector. It is very difficult for companies in the oil and gas sector to comply with the GHG reduction whereas for companies from the IT sector it is easier to reduce GHG emissions and to be a low-carbon company ... What can be very valuable for my government and other analysts is to get more sectorised information, including rankings ..., maybe that is what is missing in CDP data"* [Government 2]. In addition, NGOs would like to see questions related to social impacts produced by environmental issues. At the moment, NGOs are not able to understand the impact of environmental issues beyond companies' decisions and financial risks.

Finally, with regard to stakeholder relationships, academics, investors and NGOs mentioned the important role CDP plays in engaging stakeholders. However they demand a more active collaboration with CDP. Academics especially would like to see universities working together with CDP, using CDP data in relevant research projects that impact both academia and the market. They suggest the creation of an academic community inside CDP in order to facilitate a more active collaboration, providing a space to promote academic publications and show how they are being used. *"The idea of an academic community is good, to see what other people are doing with the CDP data. Having like a small directory interested in presenting ideas about how CDP is being used"* [Academic 2].

5.4.4 Impacts

Interview results reveal that the impact of CDP adoption differs by stakeholder group. The majority of the companies indicated that CDP does create an impact on their environmental strategies by enhancing their knowledge on environmental matters with impact on businesses, catalysing the interest from the market. In addition, their participation in CDP increases their reputation. Investors recognised that CDP brings them better quality data, which allows them to analyse financial risks dependent on environmental aspects in companies' evaluation. Governments recognise that CDP is a useful tool to formulate environmental policies at the national and international level. They see CDP as part of the solution to combat climate change and build low-carbon economies, helping them to take decisive action. From the perspective of

the academics, there is a clear use in their research and teaching, although they demand more collaboration from CDP. NGOs think that CDP has encouraged them to raise awareness in the private sector of environmental issues with a human impact. Finally, software/data providers consider that CDP is bringing new opportunities to the market in order to develop sustainability products with more value for customers. However, they also pointed out that more work needs to be done by CDP in order to facilitate the use and demonstrate the full value of its data.

5.5 Discussion

The results offer three different levels of learning to discuss. Firstly, to what extent do the results contribute to Elliot's model regarding the role of IS to improve the environment? Secondly, how do these results align with the literature? And thirdly, what can CDP learn to increase its influence and impact on companies' and stakeholders' decisions?

5.5.1 Contribution to Elliot's model

Elliot's model presents a social-technical system in which IS serves as a central element between stakeholders and certain activities to address environmental concerns. We used this model to evaluate how business transformation can be boost according to the results of our study.

The conclusions suggest that in order to carry out the sharing, understanding and integrating activities of the model, the following actions should be put in place:

- **Shared understanding:** Providing a space for discussion, offering a hub of knowledge and expertise in environmental matters;
- **Integrated activities:** Encouraging disclosure on environmental aspects in businesses and promoting the use of its data to a diverse group of stakeholders.

Regarding the initial (and unlimited) stakeholders identified by Elliot, we agree on the relevance of companies, society, governments and organisations to reduce environmental impacts. However, we discovered that investors and academics also have a role to play as contributors to positive environmental change.

Our results reveal which particular IS solutions are required, considering the constraints, improvements and adverse impacts found:

- Provision of high-quality information on a more frequent basis, improving data formats, data completeness, data consistency and reliability.
- Data accessibility, offering solutions to facilitate the data consumption with internal processes and external applications.
- Data interoperability capabilities within financial and non-financial datasets, supporting more integration with other sustainability initiatives and corporate datasets, to enable cross-checks, contextualize and complement certain information. This feature is explored in more detail in chapter 6.
- Analysis and visualisation, in order to make the comprehension of the data easier, enabling a better understanding and analysis of information considering different stakeholder interests.

As a contribution to the model, we believe that properties to enhance the internal management and decision-making processes are required to increase the adoption of environmental actions. These properties can be encouraged if the following recommendations are taken into account:

- **Reduce costs:** Organisations recognise the growing importance of taking environmental matters into consider while making their decisions. However, the usefulness of environmental information is considered relevant only on an *ad hoc* basis.
- **Make environmental decisions more relevant to other areas:** Companies recognised the difficulty of making their environmental strategies relevant within the organisation. More solutions are demanded to engage key areas within the organisations, in order to allow more integrated environmental assessment and management.
- **More data coverage:** The CDP data is recognised as allowing a proper understanding of the environmental impact on business decisions, but more frequent sector-specific data linked with financial implications and human impact matters are demanded.
- **More active collaboration:** Stakeholders are willing to collaborate more actively with reporting initiatives such as CDP and explore the new possibilities and opportunities offered by its dataset in diverse areas.

5.5.2 Contrasts with previous findings on CDP literature

The results are partly in accord with some of the conclusions from the literature regarding the role of ICT in environmental sustainability, namely with respect to constraints on data quality, reliability of data, and the lack of adequate IS for engaging stakeholders with this information, as is mentioned below by some authors:

- Lack of IS to capture and manage environmental information [76]. Most of the participants agree with this statement when dealing with CDP data. They need to transform, clean and load the data into local applications or analytical processes. There is an apparent lack of integration of environmental information within internal business processes, as disclosers and users of the information.
- Better data quality and greater reliability [58][63]. The users of the CDP data demand better quality data regarding completeness, format, structure and consistency. They find the CDP data difficult to analyse and integrate within internal processes, and highlight that too much time is required to carry out a proper analysis. Some claim the problem is not the content but the format.
- Lack of verification and audit processes [58]. Investors and academics mention that in order to consider CDP data in relevant decisions and research, they need reliable data. More verification and audit processes on environmental information are required, just as they are required when analysing financial data.

Some interview findings are at odds with the literature. Looking at the drivers and impacts of the different stakeholders, these are the contradictory statements:

- The absence of compulsory reporting suggests that environmental matters are not a priority in organisations and, therefore, it is not effectively managed [65]. The results show that companies consider environmental aspects as part of their strategies. They voluntarily report to CDP due to the perception of creating value by differentiating themselves in the market and the possibility of enhancing their business opportunities. There is also interest in

strengthening their communications with stakeholders, and in learning from their competitors and improving their reputation.

- The pressure of being targeted by a group of investors is what enhances the effectiveness of promoting changes within the organisation according to Song and Szewczyk [60]. The results partly agree with this statement. However, companies consider stakeholders other than investors, such as their competitors and the general public. Internal strategies, an improved reputation and strengthening stakeholder communications to increase business opportunities are additional factors that motivate them to change.
- CDP has a unique concern for investor interests, showing a lack of real benefit of this information for the general public [58]. The drivers and impacts identified per stakeholder group clearly demonstrate that there is an individual interest in CDP data and this goes well beyond investors. In this respect, CDP creates some advantages that no other initiatives or data repositories are offering at the moment.

5.5.3 What can be learned from CDP

From an industry perspective, the results demonstrate that the information centralised by CDP represents a substantial resource of information for the decision-making processes in companies and their stakeholders. However, CDP needs some additional work to reach the full potential of its data in such processes and influence.

The level of relevance of each property differs per participant group, as is presented in Figure 15. In general, interoperability and data quality are the most frequent demands identified. The reason is that these properties are indispensable for a proper analysis which, in general, impacts on decision-making related to environmental matters. In decisions where companies are involved, CDP data is usually considered as an additional resource to study with others environmental, social, financial and governance sources of information. As such, interoperability is identified as important to enable cross-checks, contextualise and complement certain information.

Data coverage and data accessibility are common demands, meaning that, in general, the CDP audience asks for more data on a more frequent basis and with better access mechanisms. The general trend is that companies and stakeholders perform analysis using internal tools and techniques. Most of them look at specific CDP data points, being monthly and quarterly, the normal frequency of access. In terms of data coverage, investors and data/software providers do not demand that property; their attitude is more orientated to make use of what data is already on the market, and build methodologies and products at the top to cover certain demands.

Better tools for analysis and visualisation are in demand, the main reason for this is that stakeholders prefer to use the internal applications and processes that they already have in place to perform analysis with the CDP data, instead of using those offered by CDP. This general trend does not apply to either companies or data/software providers. In the case of companies, they find these tools useful during the time they are completing the CDP questionnaires, and afterwards, when they check the results against their competitors.

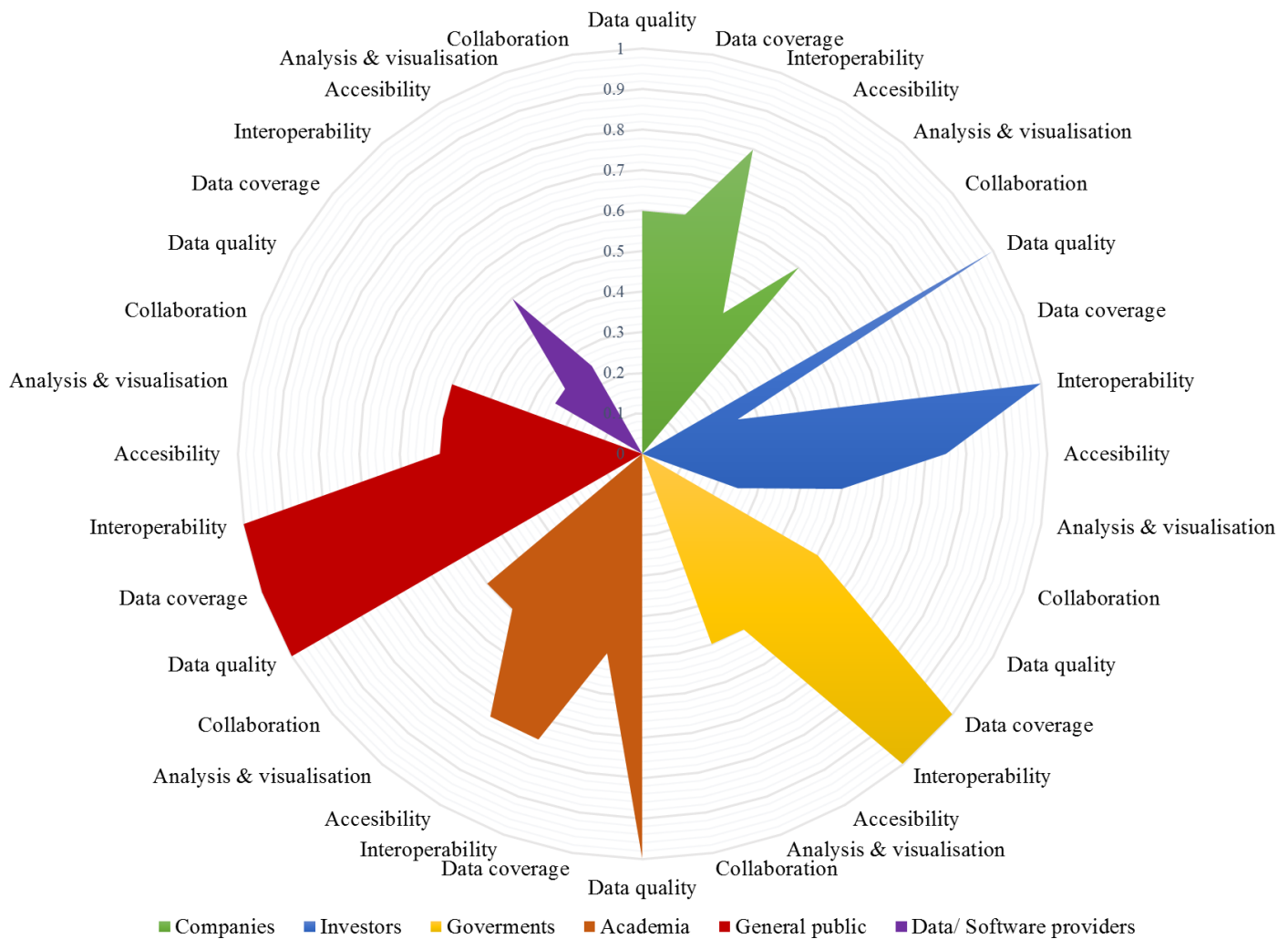


Figure 15. Level of reporting demands per market

5.6 Conclusions

The main goal of this chapter is to clarify the role of IS in the field of environmental sustainability.

During this study, we evaluated why and how relevant actors use environmental information, what their decision-making processes are, what information needs they have, what they might be in the future and what is necessary to enable this. We interviewed companies, investors,

governments, universities, software/data providers and NGOs that are CDP members and users of environmental information.

The interviews identified that the main motivations that encourage companies to disclose and use environmental information from CDP are the inclusion of better environmental strategies in their decisions, to strengthen their stakeholder communications and to enhance their reputation. In the case of investors, their main driver is to get improved accuracy on financial risks assessment, with a clear impact on sustainability methodologies for company evaluation. Research on environmental accountability, a company's environmental profile and financial impacts are the main areas of interest for academics. NGOs normally use the environmental information provided by CDP to increase the awareness of environmental stewardship in society and businesses, with a clear impact on stakeholder communications. The main driver for software/data providers is to enhance the value of the risk and sustainability assessment products and services they market to companies and investors. Despite the different drivers and impacts, most of the constraints and improvements identified are shared by all stakeholders.

Three types of learning arise from this study, contributing to the literature and promoting further developments in the IS industry and CDP.

The role of CDP is demonstrably important for reducing environmental degradation, enhancing stakeholder communications, providing a space for discussion and collaboration, and offering a hub of knowledge and expertise in environmental matters. What makes CDP different is its expertise on environmental matters, its unique dataset, and the relevance of that dataset to a diverse group of stakeholders. Nevertheless, in order to achieve its mission, CDP needs to facilitate the use and understanding of its data, taking into consideration the six socio-technical properties referenced in the results section. Interoperability and data quality are the properties most in demand. The majority of the participants consider the CDP data sets as complementary to other environmental, social, financial and governance information.

As a main conclusion, this study demonstrates that in order to achieve business transformation, three things are required:

- (1) More active collaboration between stakeholders and sustainable initiatives like CDP;
- (2) More data coverage: bringing access to more environmental content;
- (3) Innovative IS solutions to deal with environmental management and information in the areas of data quality, data interoperability, data accessibility, analysis and visualisation.

The results of this study guided the main areas to explore throughout the remainder of this research. Within the following chapters we address some of the challenges highlighted, such as: interoperability in Chapter 6, data standardisation in Chapter 7, accessibility and analysis in Chapter 8 and more data coverage and innovative solutions to deal with data quality in Chapter 9.

Chapter 6. The role of interoperability in sustainability decisions

Given the results of Chapter 5, we identify (1) ‘interoperability’ as a critical component for business transformation and decision making in complex information scenarios, and (2) CDP as a mechanism that supports environmental impact-aware decision-making processes for industries, corporations and organisations. This chapter is trying to answer our first objective “Market understanding on the use and disclosure of environmental data”.

In this chapter, we explore the role that interoperability can play in achieving business transformation in light of the European Framework for Interoperability (European Commission, 2015). In this context, interoperability is understood as the ability to exchange and make use of information between organisations and data sources. Through interviews with seventeen CDP members, we identified that the information held by CDP is seen by its members as useful for ‘environmentally sustainable’ decision making, i.e. decisions based on both financial and non-financial factors, but its usefulness is challenged by interoperability issues on several levels.

The most important results from the study were:

- Trends, demands and the place of environmental concerns in decision-making processes.
- A number of properties relevant to interoperability are demanded by CDP clients to consider environmental information in decisions.
- The CDP has a role to play in environmental sustainability decisions.

The work in this chapter was previously presented and published at the 4th International Conference on ICT for Sustainability (ICT4S 2016) ¹⁷.

¹⁷ 4th International Conference on ICT for Sustainability (ICT4S 2016) ¹⁷and published by Atlantis Press.
doi:10.2991/ict4s-16.2016.26

6.1 Introduction

The decisions made in the course of business—and the subsequent actions carried out by companies as a result—can have a negative impact on the environment. In order to solve these environmental challenges, it is necessary for companies to look beyond financial aspects when making decisions. For this to occur, businesses need to transform how data is used to inform their management and operations. The information is critical to ensure the decision-making process is based on sound evidence that takes all factors into consideration.

Such a transformation requires a deep understanding of the current actual, and potential future, environmental performance of a business. This in turn requires that firms provide better quality data and actions, and, together with their stakeholders, develop more sophisticated decision-making processes and associated information systems (IS) to support the transformation towards sustainable developments.

Several authors point out the role of interoperability as an enabler of sustainable decisions [129][130][131]. Interoperability allows firms to effectively share and reuse existing information among interested parties by:

- the exchange and use of information between information systems,
- the definition of relationships between different data environments,
- the coordination of business processes among the various organisations, and
- the coordination of different legal frameworks to allow interoperation.

The objective of this chapter is to explore the role of interoperability in environmental sustainability at present, look towards how it can improve in the future, and identify what is necessary to enable this to occur. The discussion in this chapter is based on the interviews carried out with CDP stakeholders, as described in Chapter 5.

6.2 Background

6.2.1 Standardisation and interoperability

The corporate reporting ecosystem consists of a scenario in which firms and other interested parties use the available information to make better decisions. Information and Communication Technologies (ICT) are essential in this ecosystem, and therefore should support firms' decision-making processes to allow for the required business transformation that would help to reduce environmental degradation.

Several authors advocate that both standardisation and interoperability are indispensable to ensure the effective interpretation, exchange and use of the information among different users [102][103]. In this study, both are considered.

Standardisation is key to facilitate the exchange of information, formalising the technical requirements to ensure the quality of the information. In order to understand the relevance of standardisation, the particular case of XBRL (eXtensible Business Reporting Language) is worth considering. XBRL is a standard technology required by regulators and supervisory agencies across the globe to gather financial information from large corporations, SMEs and public administration. The use of XBRL offers the following benefits [104][105]:

- XBRL offers the possibility to represent business facts, which provides data contextualised under business requirements (presentation, period, legal references, calculations) and data quality.
- XBRL enables a good level of interpretation, given the detail of the data represented.
- XBRL allows for easier accessibility and integration of the information to any application or management process, as it is an open standard.
- XBRL enables the validation and comparability of information.

The main idea behind XBRL is standardisation. For a specific rule or guideline i.e., IFRS, US-GAAP, GRI, EMAS, CDP, etc. and its corresponding statements and reports, a single XBRL taxonomy is created. The taxonomy is where the rules and data definition materialise, it is comprised by a set

of elements i.e., KPIs and all the presentation, calculation and logic rules that operate, according to that rule or standard. Once created, the XBRL taxonomy is made public as an open remote resource on the Internet. Then, for a specific firm, the proprietary software can create an XBRL instance the report itself, containing the concrete facts and figures for a certain period. The XBRL instance can be checked against the taxonomy by all parties reporting entity, a regulator, or even the public in order to guarantee its validation. The creation of an XBRL taxonomy implies the agreement of all interested parties. When applied to financial information, the XBRL working groups involve regulators, IT experts, the academia and the industry; this was also the case for non-financial guidelines like GRI and CDP.

However, standardisation is not enough to ensure effective interpretation, exchange and use of information, between people, organizations, processes and systems in the corporate reporting ecosystem. As this data scenario implies a greater level of complexity, generated by the heterogeneity of the organisations and information involved, different information content, data formats, data frequencies and ways of access, it brings to light additional challenges that require interoperability beyond simply a shared standard.

6.2.2 The case of CDP, an environmental reporting initiative and a source of environmental data

As we mentioned in previous chapters, CDP now holds the world's largest collection of information on climate change and the impact of corporations on water and forests. The aim of this data is to support investors in better understanding associated risk and encouraging long-term business transformation towards sustainability.

CDP as an organisation is active in increasing the consistency and alignment of its data with other reporting organisations and frameworks. This organisational cooperation takes the form of MoUs and technical documents, which detail the levels of alignment between CDP data and other

reporting frameworks. Currently, CDP maintains collaboration lines to align areas of reporting with organisations such as the Global Reporting Initiative (GRI) and Dow Jones Sustainability Index (DJSI). The main goal is to enable users and reporting firms to refer to the same data points through different reporting channels. It represents a significant step towards the global standardisation of environmental reporting.

Although this dataset is useful to a multitude of companies and other stakeholders, it is not yet clear what the real use and impact of this data is on decision makers. Questions arise regarding the place of CDP as an environmental data provider in the corporate reporting ecosystem and its impact on decision makers. Likewise, considering the role that CDP plays, it is unknown how interoperability can bring new opportunities to drive forward business transformation. We consider CDP relevant for this study because:

- It offers a framework to disclose environmental information to companies and cities from around the world, effectively driving information standardisation;
- It has a voluntary reporting system which centralises standardised environmental information;
- It provides its data to decision makers in the expectation that they use it to inform their actions;
- It is an environmental reporting initiative which puts environmental information at the heart of financial decisions making, to help reduce environmental degradation.
- It is a reporting initiative [56][64][67] and a source of data explored in the literature [58] [61] [63].

6.3 Research questions

This particular study explores how interoperability currently supports stakeholder decisions towards making sustainable actions, and how can it enhance such decisions in the future. This

investigation took place as part of the qualitative study presented in Chapter 5, and shares the same methodology but only 17 participants responded to the interoperability aspects.

Using the same methodology as the previous study. We categorised the results of different interoperability properties, evaluating the attitude of the participants towards interoperability, considering the need and current practices mentioned by each participant and addressing the four layers of interoperability (legal, organisational, semantic and technical). As a result, a set of interoperability properties demanded or adopted by the participants to include environmental information in their decision-making processes were identified. We consider an interoperability property as any feature needed to address any layer of interoperability. The set of properties is presented in Table 4, classified under the correspondent interoperability layer and the participant who made the comment.

The results provide a set of interoperability properties that enable us to measure the attitude of companies and their stakeholders towards interoperability, and in turn determine how relevant each layer is. During the next section, we explain our interview findings with regard to each interoperability property.

6.4 Results

As our objective is to explore the role of interoperability in stakeholder analysis and decisions, we identified the ways in which the interviewees used the environmental information provided by CDP.

In summary, these are the interoperability trends we identified and which will be explored in more detail in the following sections:

		Levels of interoperability										
		Policy		Organisational		Semantic		Technical				
		Policy integration	Policy compliance	Link CDP information with others global standards: GRI,WRI,MSI, IPCC. GHG protocol...	Link environmental aspects between different areas of companies	Link CDP data with other sources of corporate information: Annual reports, stock market data	Link different CDP datasets (climate change, water, forest)	Data format	Data connectivity	Data accessibility	Data frequency	Data integration
Companies	Company 1			✓		✓	✓		✓			
	Company 2		✓	✓	✓	✓	✓	✓	✓	✓		
	Company 3			✓		✓	✓	✓	✓			✓
	Company 4		✓	✓		✓	✓	✓	✓			
	Company 5			✓	✓			✓		✓		
Investors	Investor 1	✓		✓		✓	✓	✓	✓			✓
	Investor 2			✓		✓	✓	✓	✓	✓		✓
	Investor 3	✓		✓		✓		✓	✓	✓		✓
	Investor 4	✓		✓		✓	✓	✓	✓	✓		✓
Governments	Government 1			✓		✓	✓		✓	✓	✓	✓
	Government 2	✓		✓		✓	✓	✓	✓	✓	✓	✓
Academia	Academic 1			✓		✓	✓	✓	✓	✓		✓
	Academic 2	✓		✓		✓		✓	✓	✓		✓
	Academic 3	✓		✓		✓	✓	✓	✓			
	Academic 4	✓				✓	✓	✓	✓	✓	✓	✓
NGOs	NGO 1			✓				✓	✓	✓		
	NGO 2							✓		✓		

Table 4. Interview results and interoperability properties

6.4.1 Legal interoperability

The analysis of the results suggests that the participants consider two aspects of legal interoperability to be relevant.

- **Policy integration.** The concept of interoperability in policy is defined by the level of compatibility with datasets evaluated for decisions. Some companies, investors and academics mentioned that they evaluate legal conditions in order to determine the level of compatibility with the current datasets consulted for their decisions, as policies can demand new information flows. These may be relevant to integrate within their analysis. [Government 2]: *“Another analysis we do is tracking what is being done at the company corporate level and how it links to the national policy level. We are interested in evaluating the impacts of company actions on traceability at the national level”*. It is mainly investors, governments and academics who consider it pertinent to assess how the companies that respond to CDP comply with regulations at national and global levels in order to determine how confident these companies are with their current direction. [Investor 1] *“As part of our company’s evaluation criteria, we analyse the level of commitments with governments and if they are ready for future regulations”*.
- **Policy compliance.** For companies, the policies have an impact on the environmental actions they take and the information they need to disclose to certain authorities. It affects their reporting processes, which results in more data to store and more information to communicate. [Company 4] *“Reporting is not a standalone thing, and we have constant reporting influences with impact on operations. These actions that we are taking, we talk and report about. We also have to comply with legal regulations, so we need to feed into that channel as well”*.

6.4.2 Organisational interoperability

With regards to organisational interoperability, there are two levels of coordination demanded by the participants: (1) greater focus on the convergence of CDP data with other reporting standards, and (2) closer working practices of CDP with companies to help them align environmental strategies within their departments.

- **Link CDP information with other global standards.** The majority of the participants agree that CDP should align strategies and content with other reporting organisations, specifically with the most recognised sustainability frameworks in the market. From the companies' perspective, this would represent an improvement during their disclosure process. By aligning the different frameworks used to support their corporate reporting, it would result in a reduction of effort because the same information would no longer need to be reported in different ways. [Company 3] *"We do not only disclosure to CDP, also we use GRI framework to generate our sustainability report. There are certain indicators from the environmental section of GRI aligned with CDP, however, the way the information is asked, in term of level of aggregation, units and methodologies are completely different. So we need to make a double effort to report sometimes the same information in different ways"*. In order to facilitate improved validation by allowing cross-checks between common data points, data users demand that CDP aligns with other major frameworks by sharing areas of interest and content. This is needed because companies criticise the absence of verification and audit processes in the CDP data. They also mentioned there is a lack of technical agreement between organisations regarding data formats, content, level of aggregation and unique company identifiers. These features are needed to facilitate the consumption and combination of data from multiple frameworks, and to get more complete information for decisions. [Investor 1] *"Ideally, we would like to integrate CDP responses with GRI, WRI and SASB information available, and evaluate the overlay of both responses and results. That would be useful to consider non-financial information in our most relevant decisions. Currently, we cannot do that level of analysis, because each dataset is published with different formats,*

structures, locations and content. Moreover, it requires a manual process to find levels of alignments to analyse."

- **Link environmental aspects within different areas of companies.** Companies find it difficult to engage key departments within their organisations regarding environmental concerns to get a more integrated environmental assessment and management to disclose to CDP. The interoperability of environmental information and alignment of environmental reporting within different areas of companies is needed, but the lack of common environmental interests and concerns is presented as the main barrier. [Company 5] *"We found CDP useful because it helps us to understand the value of considering certain environmental aspects of our business. However, it is still difficult to cover all the points demanded by CDP due to the dependencies with other departments of my company, which are not aligned with that level of environmental concerns and, therefore, management".*

6.4.3 Semantic interoperability

After identifying the legal and organisational mechanisms demanded by the participants to make better use of CDP data in their analysis and decisions, there are two main semantic requirements:

- **Link CDP data with other corporate datasets.** With the exception of NGOs, the majority of the participants would like CDP to facilitate the connection of its dataset with other publicly available corporate information, such as financial reports, stock market data, corporate website and social media profiles. These types of data sources are the most useful to carry out analysis and contextualise findings for these businesses. [Government 2] *"I think what is really important for us is to contextualise the CDP data. To tell a nice story is very useful for us... for example, demonstrating that there are numbers of companies taking actions representing the revenue at a certain amount. That provides very powerful statements...But to be honest, it is difficult for us, as we have to put together certain company data such as revenues, total of employees, assets in management, which is difficult and time-consuming. If CDP can provide*

that information, it would be very helpful". However, the lack of alignment in data formats, unique identifiers, data frequencies and the lack of data accessibility are mentioned as the principal difficulties in carrying out deep analysis with multiples datasets. Connecting pieces of information about the same company from different data sources requires semantic alignment to take place.

- More integration across the different CDP datasets.** Publishing the climate change, water, forest and scores datasets integrated into one is regarded as an important improvement for CDP by companies, investors, governments and academics. As potential users of the information, they expect CDP to get relevant information about a company's environmental profile. However, CDP deals with each dataset separately, collecting the information via three different questionnaires without linked information. This means that CDP does not offer an integrated environmental profile where, for instance, a company's risks can be evaluated by water, climate change and forest aspects. In order to do that, stakeholders need to work with each dataset separately and to consider factors such as companies not necessarily responding to the three questionnaires. [Investor 1] *"Would like to see more integration of CDP datasets. At the moment, water, forest and climate change appear separately, which is not very useful for us when we need to evaluate the environmental profile of any company"*. Table 5 describes the external data sources identified and the reason why the link with CDP data is relevant.

External data source	Why is the link with CDP data relevant for?
Sustainability frameworks and indexes, such as UNFCCC, IPCC, GRI, and DJSI	To cross-check data points and complement analysis and strategies, improving the accountability of sustainability indicators.
Corporate reports, such as governance reports, annual reports, CSR reports, risk reports	To cross-check specific data points and evaluate the level of integration of FESG topics within companies

Financial and stock market data	To evaluate specific impacts of ESG topics in financial markets
Social media data	To identify key topics for the society and communicate externally
Regional and local policies	To keep updated about future regulations in CSR reporting areas

Table 5. External sources to link with CDP data

6.4.4 Technical interoperability

As part of the interviews, several technical needs were mentioned as requirements to reach other levels of interoperability:

- **Data format.** Most of the participants find it difficult to carry out deep analysis combining CDP data with others datasets, comparing multiple companies and data points. [Academic 3] *"When I look at a company's response in PDF format, I know it is typically 40 pages. The qualitative data is very difficult to read, it is very difficult to copy and put in a different format as the formatting is locked. ... Probably my biggest criticisms is not the content, it is the format of the content."* *"At the moment CDP just offer a flat dataset and a confused spreadsheet which are not being used in a way which is intelligent."*
- **Data connectivity.** The participants are interested in having technical solutions to help them consume different types of company information. For example, they would like to know how to link CDP responses with other sources of information which refer to same companies, like stock market data and sustainability reports. In that way, they want to add more value to their analysis by contextualising financial and environmental impacts. Also, they demand technical facilities to find common data points that CDP shares with other reporting frameworks like GRI. This means that if a company discloses their carbon emissions in their annual report and also to CDP, they can have the capacity to cross-check both sources of information for validation purposes. Some academics pointed out that the goal of CDP is not to provide

datasets for financial research. However, CDP's audience demands the match between CDP data and financial information. The main challenges recognised are the lack of harmonisation between datasets and unique identifiers in corporations. [Investor 3] *"Contextual information is key when analyzing companies performance. For us, CDP data is never analyzed alone, we always crossed that information with specific financial information. If that link to other data sources is given by CDP, it will save us a lot of time due to it is a very time-consuming task"*.

- **Data accessibility.** As mentioned in Chapter 5, accessing CDP data is complicated and difficult. Academics, governments, investors and companies find it necessary to have better mechanisms for data consumption and analysis.
- **Data frequency.** Some investors and academics, and most governments, consider it a constraint to have the CDP data on an annual basis. For their progress analysis and assessments, they would like to have data on a quarterly or monthly basis, aligned to financial obligations, in order to make more accurate decisions. [Government 1] *"Data on an annual basis is not enough to show the momentum. We need data to check company and city progress and commitments over time. That is the way to see what it is going on and how things are going to occur "*.
- **Integration with existing data platforms:** Investors, academics, governments found it useful that CDP information is integrated in data platforms like NAZCA, Bloomberg, Sustainalytics, Thomson and Reuters. [Investor 1] *"As CDP data can be accessed via a Bloomberg terminal, this is one of the reasons why it is useful for us. Because if I am in Bloomberg, I do not want to go off and go to other website and log in, to pull all the data. That is the reason why it is so useful. Just thinking that having more and more information in what future companies are disclosing, makes us think in more sophisticated ways to measure and monitor in order to quantify some of the ESG issues. We will have the need to have more datasets in one place"*. However, in some cases, the CDP data integrated into these platforms is not enough and does not cover stakeholders' demands. [Investor 3] *"In Sustainalytics, there is data related to CDP,*

whether the company respond to CDP or not. However, we do not have immediate access to what the company responds. In that case, we need to access the data through CDP”.

6.5 Discussion

The results demonstrate that interoperability has a role to play in the development of decision-making processes and associated IS, where particular social-technical properties under the influence of general developments in ICT are required to drive long-term decisions, with impact on strategy and management processes. A major level of technical formalisation is required in terms of convergence and applicability of environmental initiatives like CDP with other reporting organisations, data providers and areas of companies. Moreover, it is necessary to have ICT solutions to exchange data in a meaningful manner and make the data more accessible to a broader community by means of a major harmonisation and integration with external applications and organisational processes. The results show that in order to achieve a major harmonisation and integration, aspects such as data standardisation, data connectivity, data accessibility and data integration with external applications must be considered. These results contribute to ICT for sustainability and sustainable decision support systems studies [106][107][108], while also identifying concrete information systems properties required by real users of environmental information.

Looking at the context where stakeholders use CDP data, we have discovered that interoperability is implicitly considered in policy, organisational, semantic and technical levels. Likewise, we found that CDP has further work to do to allow users to unlock the full potential of its information, specifically:

- **A major policy engagement.** Policy represents more environmental information for companies to store and disclose, and in turn more data for their stakeholders to evaluate in order to make decisions. If CDP increases its level of interoperability with forthcoming regulations, defining what its level of compatibility and impact is, it will help reporting

companies during their disclosure process and also their stakeholders in making better decisions. Better policy engagement with forthcoming regulations brings opportunity to CDP in order to increase the value and impact of its data.

- **More detailed organisational agreements.** CDP should agree and define with greater detail the level of alignment of its data with other frameworks. MoUs and current technical specifications are not enough. It is necessary to concretely define details about data formats, content and scope, transformations and the levels of aggregation. In that way, companies would not need to report the same information in different ways, depending on the final report (e.g. whether a CDP report, annual report, sustainability report, etc.). Being more specific about this alignment allows CDP data to be considered more credible and reliable, facilitating cross-validation with external data sources. CDP also should work more closely with companies to help them improve engagement on environmental aspects within their different departments. Better integration and management of environmental factors in companies would enhance environmental data quality. However, that represents an additional effort for CDP, given the variety of international companies that already disclose to CDP.
- **Enable better semantic interoperability.** The need for using and analysing CDP data in relation with other datasets from financial, social and environmental topics is a common demand from the majority of companies and their stakeholders. The reality is that combining CDP data with other corporate datasets or corporate reports is difficult and almost impossible. The main problems are the lack of harmonisation between the different datasets regarding data formats, content, data frequencies and unique identifiers. One of the solutions CDP can offer is to provide its data in different data formats based on how other data sources are presented and which formats are familiar to the CDP audience. For example, CDP could provide its data openly and in open standard formats, such as XBRL, as some of the reporting companies

disclose their financial information to regulators like the U.S Security Exchange Commission¹⁸ in XBRL format and investors are familiar with analysing XBRL information for their decisions. Furthermore, CDP could potentially offer companies the possibility of submitting some of its data related to targets on a more frequent basis, which will help use the information for monitoring purposes. Finally, CDP should make its data linked to other sources of information accessible or present its data with a company's unique identifiers and teach the audience how to use it in combinations with other sources of information. Overall, semantic interoperability in CDP will help improve the decision-making process so that it includes social and environmental aspects as well as financial information.

- **Better technical solutions.** There are three types of technological aspects demanded of CDP. First, to introduce standard formats to facilitate the exchange and combination of CDP data with other sources of information. The need to standardise data in CDP brings up the possibility of adopting XBRL for non-financial data. Also, publishing CDP data linked to other structured data, using best practices in semantic web such as linked data¹⁹, could be a potential solution to explore. Secondly, CDP should provide access to its data automatically, enabling better integration with external applications and internal processes. Ideally, CDP should include the ability to retrieve and manipulate its data for greater accessibility. Possible solutions include those based on an Application Programme Interface (API), web services or a semantic interface for querying, such as SPARQL²⁰, which enables users to write queries to get access to the data and its links to other data sources. Ideas that have been implemented are discussed in Chapter 8. Finally, it would be beneficial to have better integration of CDP data with external data platforms, such as NAZCA, Bloomberg and Sustainalytics, in order to add value to decisions. For this to occur, CDP should work more closely with data providers to transmit the full value of the data. CDP should be more active in encouraging data providers to present and use CDP data through external platforms.

¹⁸ U.S Security Exchange Commission, 2015 <http://sec.gov> [Accessed August 2019]

¹⁹ Linked data, 2015 <http://linkeddata.org/> [Accessed August 2019]

²⁰ W3C, 2008 <https://goo.gl/8KvJJW> [Accessed August 2019]

6.6 Conclusions

We explored the role of interoperability in sustainable decisions, using CDP as a case study. We conducted interviews with a set of companies and their stakeholders, as we did in chapter 5. As a result, the following three learnings arise:

- (1) Current reporting trends and demands from companies and their stakeholders, and the place of environmental concerns in decision-making processes.
- (2) A set of interoperability properties are needed to allow for environmental information to be considered in decisions.
- (3) While CDP has a role to play in environmental sustainability decisions.

Companies and their stakeholders demonstrate that their reporting trends are evolving to take advantage of the sources of information that corporate reporting ecosystem provides. Their decision-making and management processes consider financial, social and environmental aspects from corporate reports, other sources of information and data platforms. However, challenges arise in order to carry out a better use of the information available for making decisions and achieving business transformation. Interoperability properties need to be considered to achieve the transformation, and we demonstrate that legal, organisational, semantic and technical interoperability are deemed to increase the value of environmental information for decisions. As a contribution to ICT for sustainability studies, the need for ICT solutions to overcome these interoperability demands are highlighted in the areas of data standardization, data connectivity, data accessibility and data integration. These areas are further explored in Chapters 7, 8 and 9.

In summary, the study suggests that CDP should increase its level of interoperability to be in alignment with future regulations, other reporting organisations and respective frameworks, as well as better define its links with other sources. CDP should also provide for integration within its different datasets (climate change, water and forest) and introduce the ICT solutions mentioned above to address these interoperability demands.

The next chapter explores the role that certain technologies can play given the interoperability demands necessary in order to support decisions, such as XBRL for standardization and semantic meaning. For example, what value can XBRL bring to environmental reporting initiatives like CDP, considering the role that these initiatives play in companies and their influence on decision making and actions (Chapter 7) Additionally, given the corporate reporting ecosystem and the interoperability demands from companies and their stakeholders, it is vital to identify the main opportunities and barriers facing business transformation. This is explored in detail with solutions like XBRL in combination with linked data (Chapter 8).

Chapter 7. Influencing the practices of XBRL within CDP

This chapter is based on the results of Chapter 5 and 6 regarding the need for improving data standardisation and interoperability in the use of environmental information in the decision-making process. This chapter is trying to answer our second objective “How to make environmental data more accessible for users?”.

In particular, this chapter presents the work that I have done to influence good practices on data standardisation through XBRL technology within CDP, considering the relevant role that it could play in environmental-related reporting. My influential work consisted of three deliverables.

1. **An article published at CDP** where XBRL was presented as part of the solution to facilitate more streamlined and standardised corporate reporting on environmental variables. This work was done for CDP and i2Investing Initiative under a project called “Sustainable Energy Investment Metrics”. It was written by me and received the recognition of the following organisations: XBRL International, XBRL Asia, Eurofiling, the European Committee for Standardization (CEN), the Spanish Accounting and Business Administration Association (AECA). It was published in the context of H2020 and funded by the European Union’s Horizon2020 research and innovation programme under grant agreement No 649982. The complete article is presented in section 7.1 and can be found at the CDP website ²¹.
2. **The 2016 Climate Change Taxonomy**, which was developed only by me, it is the digital representation in XBRL of the 2016 Climate Change questionnaire. This work has been recognised by The Spanish Association of Accountants (AECA) to define the environmental data as part of their Integrated Reporting framework, which is facilitating Spanish companies affected by the European non-financial Directive (2014/95/EU) to disclose their

²¹ XBRL report: <https://www.cdp.net/en/articles/climate/technology-and-climate-reporting-can-xbri-help-tcfd>

environmental. This integrated reporting framework developed by AECA is also referenced in the Spanish transposition of the European Commission law. In addition, this taxonomy is being utilized as a good example for European regulators to consider the possibility of requiring environmental data in a more standard and digital way aligned with existing financial practices. As this European directive is evolving to be required in a more normalized and structured way. This taxonomy is openly accessible at the XBRL International Taxonomy Registry²². The purpose of this work was three-fold:

- a. Improving data quality and data accessibility, and standardising Climate Change data and business rules in an open digital format.
- b. Connecting environmental and financial information models by using the same reporting technology adopted by financial information markets worldwide (e.g. IFRS, USGAAP, UKGAAP, Spanish GAAP, EDINET, etc). XBRL is already required around the world, by regulators and supervisory agencies.
- c. Improving the consistency of environmental data across other sustainability disclosure frameworks and reducing reporting burdens by aligning data contexts across other sustainability frameworks that use XBRL, such as GRI (Global Reporting Initiative) and DJSI (Dow Jones Sustainability Index).

This complete work can be found on the CDP website²³ and the technical insights and modelling decisions can be found in section 6.2.

3. Dissemination of this work and related ideas within relevant industrial and academic forums (Figure 16) and the evolution of the standard through the author's work as Chair of the Best Practice Board at XBRL International.

- a. Mora, Maria (2017). Concept and use of enumerations 1.1: The CDP taxonomy. 23th Eurofiling Wokshop (European Central Bank, Frankfurt, Germany). June 2017

²² XBRL International Taxonomy Registry: <https://taxonomies.xbrl.org/>

²³ The CDP-XBRL taxonomy: <https://www.cdp.net/en/research/xbrl>

Work developed and presented by Maria Mora. Presentation available here: <http://eurofiling.info/2017/tutorials/enumerations/>.

- b. Mora, Maria (2016). Envision climate-related reporting challenges. *Natural Capital Summit Conference* (Madrid, Spain). July 2016

Work developed and presented by Maria Mora. Presentation and information available here: <https://naturalcapitalsummit.wordpress.com/programa/>

- c. Mora, Maria (2016). Financial Stability Board on Climate-related disclosure. *18th XBRL Europe Day Conference*. (European Central Bank, Frankfurt, Germany). June 2016

Work developed and presented by Maria Mora. Presentation and information available here: <http://eurofiling.info/portal/xbrl-week-frankfurt-2016-2/>

- d. Mora, Maria (2015). Natural Capital Disclosure. *15th XBRL Europe Day Conference*. (Bank of Spain, Madrid, Spain)

Work developed and presented by Maria Mora. Presentation and information available here: <http://www.eurofiling.info/201506/index.shtml>

- e. Mora, Maria (2014). Climate Change Taxonomy Insight. *20th Eurofiling Conference*. (National Bank of Belgium, Belgium, Brussels)

Work developed and presented by Maria Mora. Presentation and information available here: <http://www.eurofiling.info/201411/index.shtml>

- f. Mora, Maria (2013). Enumerations. *18th Eurofiling Conference*. (CSSF Luxembourg, Luxembourg, Denmark)

Work developed and presented by Maria Mora. Presentation and information available here: <http://www.eurofiling.info/201312/index.shtml>

- g. 33rd World Continuous Auditing & Reporting Symposium – Rutgers University (New Jersey, USA)



Figure 16. Dissemination of XBRL work within relevant industrial forums. (European Central Bank, 2017)

Complying with the EngD requirements, this work made a significant contribution to the industry by influencing the adoption of XBRL practices within CDP and its market.

7.1 The role of XBRL in overcoming environmental-related reporting challenges

7.1.1 Introduction

With the 2015 Paris Agreement, climate change is at the forefront of international policy and business issues. For an orderly transition, businesses need to collect relevant, timely and reliable material upon which to base their business strategy. Providers of capital, insurers and regulators also need this information from businesses in a transparent, comparable, clear and verifiable structure so that they can make the best-informed investment decisions and ensure compliance.

However, regulators, investors, creditors and underwriters are faced with the difficulties of:

- 1) Accessing that information to inform their decisions;
- 2) The quality of such voluntary disclosure; and
- 3) How to interpret the disclosure.

The lack of action and decision-making to mitigate climate change is also a consequence of the following reporting gaps:

- 1) The financial significance and exposure of assets to climate change is not assessed or disclosed; this exposure is driven by local factors and granular data might be needed to assess it (e.g., asset level data);
- 2) Misalignment between the reporting of non-financial information and financial reporting, with the links between the two still not fully perceived by both preparers and users of information;
- 3) Data quality concerning incompleteness, reliability, comparability, verifiability and structure; and

- 4) Availability of information systems that facilitate the disclosure and use of combined information within the decision-making processes of companies, their stakeholders, regulators and supervisory entities.

The Financial Stability Board (FSB), at the direction of the G20 countries, has set up the Task Force on Climate-related Financial Disclosures (TCFD) to draw up recommendations on voluntary disclosure. In this paper, we present the use of eXtensible Business Reporting Language (XBRL) as a potential solution to the availability of, and access to, information. We present further recommendations for improving the quality of climate-related disclosure and increasing the volume of companies disclosing in publication for CDP. By ensuring that global climate change disclosures are consistent and standardised in a machine consumable fashion, the FSB can ensure the utility of its efforts in this field. This is based on CDP's 14 years of experience of collecting climate-related disclosures on behalf of investors representing over \$100 trillion of managed assets and with over 5,500 companies disclosing in 2015.

7.1.2 Task force on climate-related financial disclosures (TCFD)

In 2015, the FSB established the TCFD as an industry-led group to make recommendations for improving principles and practices for voluntary climate-related disclosure. The TCFD is comprised of a diverse group of experienced members to lead it, drawing from disclosure users, preparers, and market participants from a variety of industries and regions [109]. They issued their final report back to the G20 by mid-2017, thus attempting to address current reporting gaps by:

- 1) Supporting the disclosure of climate-related financial risks and opportunities;
- 2) Promoting alignment across existing disclosure regimes;
- 3) Improving the production of consistent, comparable, reliable, clear and efficient information; and

- 4) Increasing the consideration of environmental matters on decisions in the short, medium and long-term, making the use of environmental information easier.

7.1.3 The role of technology: overcoming environmental-related reporting challenges through XBRL

To overcome current climate-related reporting challenges, it is relevant to consider the role that technology, and XBRL in particular, could play.

XBRL is a standard technology used by regulators and supervisory agencies all over the world to gather financial information from large corporations, financial institutions, SMEs and public administrations. XBRL is in use within more than 60 countries at present, implemented by over 100 regulators, covering some 10 million companies worldwide. Key regulators involved in its implementation include the U.S. Securities and Exchange Commission (SEC), which adopted rules in 2008 requiring public companies and foreign private issuers to provide financial statements in XBRL, and to publish their financial statements on their corporate websites using XBRL [132]. Since then, other regulatory agencies around the world have enacted similar mandates. In Europe, XBRL is one of the recognised standards by the European Commission to address the Digital Single Market Strategy [133]. In fact, it is now required for external financial reporting by the European Banking Authority (EBA), the European Insurance and Occupational Pensions Authority (EIOPA), and the European Central Bank (ECB). XBRL is also being examined by the European Securities and Markets Authority (ESMA) to act as the European Single Electronic Format for reporting within securities markets (Figure 17). XBRL has long been adopted by the main regulators in Japan, is the reporting format in use across the Chinese public markets and is currently being implemented as an important part of the Russian Central Bank's digitalisation program. The standard is supported by a broad ecosystem of stakeholders, including hundreds of software vendors from around the world. When implemented carefully, the burden on industry can be low, although it is important to follow a range of best practices to ensure optimal outcomes in this regard.

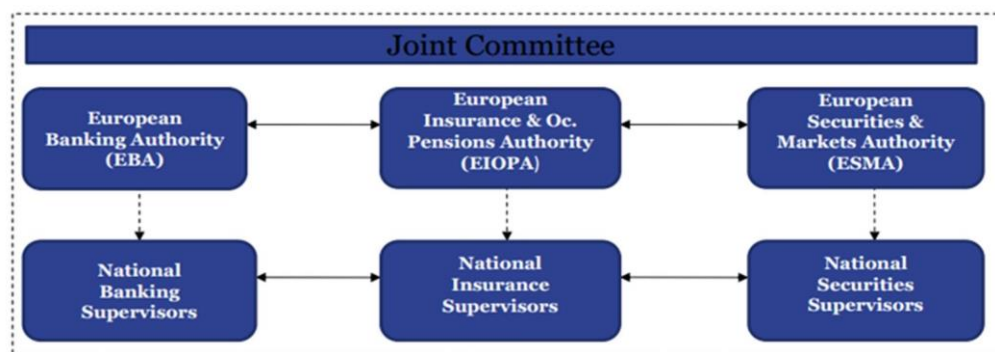


Figure 17. XBRL implementation within the European System of Financial Supervision

The standard is managed by a not-for-profit consortium, with 700 members from around the globe, with an explicit public interest purpose to enhance the accountability and transparency of global business performance through the development of open data exchange standards in this field.

7.1.3.1 What is the place of XBRL in organisations?

XBRL was born as a solution to overcome the limitations in traditional and mainly paper-based disclosures, such as ‘one size fits all’ reports, print medium fixation and one-way communication. It serves as a solution to issues in traditional reporting such as the ones highlighted by Mora and Mora [104]:

- 1) the vast amount of organisational information, both audited and unaudited;
- 2) the lack of connection between firm publications; and
- 3) the inefficiencies of a PDF-based format for report delivery.

7.1.3.2 How does XBRL work?

For financial disclosure regimes like International Financial Reporting Standards (IFRS) and US Generally Accepted Accounting Principles (US GAAP) or for nonfinancial disclosure regimes like the CDP information requests and Global Reporting Initiative’s G4 guidelines, and its corresponding statements and reports, a single XBRL taxonomy is created. The taxonomy is where the rules and data definitions are organised. It is comprised of a set of elements (i.e., Key Performance Indicators and narratives) and all the presentation, calculation and standard logic

rules that are in effect. Once created, the XBRL taxonomy is made public as an open source file on the internet. Then, for a specific firm, software can be used to create an XBRL instance (the report itself), containing the specific facts and figures for a certain period. The XBRL instance can be checked against the taxonomy by all parties (reporting entity, a regulator, or even the public) in order to guarantee its data quality and reliability, as the taxonomy contains data quality checks that any XBRL engine can validate. The validation rules supported in XBRL allow a good level of data quality, from basic rules to validate data types (number, text, precision), to more complex rules relating to elements that have been disclosed. For example, rules can be implemented to check if a breakdown of emissions is equal or not to the total emissions reported, or a CO₂ intensity figure (tCO₂/revenue) is actually in line with revenue and emissions figures reported.

The creation of an XBRL taxonomy implies the agreement of all interested parties (regulators, IT experts, academics, industry). Once the taxonomy is made public, the reporting entity must (for mandatory reporting schemes) or may (for voluntary reporting initiatives) prepare and publish the necessary XBRL reports. The report can be for multiple recipients: the corporate website, an official reporting platform, a data repository, etc. Once the data is generated in XBRL, business facts are more accessible for any kind of data analysis application, and enable all users to make quick and easy calculations, rankings, benchmarks, and comparisons. The reporting entity itself can also benefit from this digital format for management, consolidation or internal auditing purposes (Figure 18).



Figure 18. Benefits of XBRL

XBRL is used primarily for the exchange of financial, risk management information and solvency ratios²⁴ under mandatory and voluntary filing programmes. Environmental and sustainability reporting initiatives, including CDP and the Global Reporting Initiative (GRI), have initiatives promoting the disclosure and use of their data through XBRL to enhance the adoption and impact of their data for decision-making purposes [110]. Through its initiative, CDP has been working with the XBRL community to incorporate specific requirements within the XBRL standard, which are necessary for environmental reporting practices and its community. CDP regularly participates in XBRL conferences and meetings where issues of standardisation and adoption of the standard are discussed. Because XBRL provides very well understood mechanisms for defining reporting requirements in a multi-lingual context and allowing filers to use their language of choice, while allowing consumers to review it on their own, the use of the business reporting standard for this initiative seems especially relevant. XBRL International is

²⁴ XBRL has been the technical solution to implement the banking supervision regulations that have come out after the financial crisis and which imply the transmission and analysis of large volumes of data.

developing mechanisms that allow the republication of XBRL data using a wide range of technologies, including JSON.

By expanding the adoption of XBRL in financial, risk management, environmental and sustainability reporting, policy makers and industry leaders can address climate-related financial reporting challenges and drive new business opportunities in corporate reporting.

In the following sections we explain how XBRL can be a solution for better financial and environmental accountability, exchange of information and decision making when it comes to sustainability.

7.1.3.3 How can XBRL increase the perception understanding of financial significance?

Understanding the financial significance of environmental aspects depends on the ability to easily identify, in a data-driven way, their financial impacts. This requires a convergence of the current financial and environmental reporting practices adopted by corporations. For that purpose, firstly, a common agreement between financial and non-financial initiatives is necessary to identify what levels of convergences or alignments exist; and secondly, technology is required to make these agreements applicable to bring the benefits to the market. From the technology side, XBRL is the “Rosetta stone” that can facilitate this convergence, since:

- 1) it is the shared technology between financial and non-financial reporting initiatives, benefitting from a very broad ecosystem of existing XBRL capable software from large and small ERP and reporting vendors alike;
- 2) it is a proven technology that can facilitate the convergence (alignment) across frameworks, namely in terms of granular data representation and validation across countries and industries;
- 3) it is able to provide an automatic and low cost data feed to the financial data supply chain;
- 4) its value is recognized by environmental reporting initiatives like CDP and the Japanese Voluntary Environmental Reporting scheme; and

- 5) XBRL is already an adopted solution by preparers and users of corporate reporting information for their decision-makings with an established community of practice.

Bringing together financial and climate change reporting using XBRL enables preparers to integrate different corporate aspects and address the interests of a wide range of stakeholders in understanding the long-term financial, social and environmental success of companies.

7.1.3.4 How can XBRL promote alignment across existing disclosure regimes?

XBRL can help alignment between different reporting frameworks since it is able to represent relationships between different reporting models. This means that XBRL enables the generation of a coherent framework for climate-related financial disclosures using elements of information that can be found in recognised financial, risk management and environmental frameworks that use XBRL, such as IFRS and US GAAP (Accounting), Basel III (Banking), Solvency II (Insurance), CDP and GRI frameworks. This can improve the consistency of climate-related data across disclosure frameworks, reduce the reporting burden for preparers, and facilitate the data interpretation by the users. As an example, the climate exposure of a bank's lending portfolio is likely to require the detailed and granular data of its lending portfolio²⁵. This is already reported for banking supervision in Europe. The characterisation of a few more climate-relevant aspects associated with each loan, for example by the economic activities code like the NACE²⁶ codes, could facilitate the generalised assessment of the climate exposure of banks to high-carbon and high-risk projects and companies, given a common level of analysis and appropriate methodologies.

²⁵ Example of breakdowns by the European Banking Authority:

<https://www.eba.europa.eu/documents/10180/1028653/ITS+on+Supervisory+reporting.pdf/9212b4e7-37a14bbf-8409-2cc450d8513e>

²⁶ Nomenclature of Economic Activities is the European statistical classification of economic activities. For a review of those taxonomies check 2ii paper on "[Decree implementing Article 173.vi of the French Law for Energy Transition – Challenges and first recommendations](#)".

7.1.3.5 How can XBRL facilitate useful, consistent, comparable, reliable, clear, and efficient disclosure?

Standardisation is key to facilitate the exchange of information, formalising technical requirements to ensure the quality of the information. Through data standardisation in an open digital format, XBRL can help enhance data quality and data analysis through:

- 1) An open mechanism to represent contextualised business facts under defined business requirements (presentation, period, legal references, calculation) and data quality;
- 2) Enabled data-driven decision management, given the detail of the data represented;
- 3) Improved accessibility and integration of the information to any application or management process, as it is an open standard; and
- 4) Standardised validation and comparability of information.

As an example, XBRL can be used as a means to implement existing taxonomies characterising high-carbon/low-carbon products and assets⁵ which, coupled with granular revenue data, can facilitate the automatic analysis of current and future exposure to the low-carbon transition.

7.1.3.6 How can XBRL support the evolution of information systems for climate risk-related financial disclosure?

XBRL allows information to be accessible to any application for data processing and analysis for quick and easy calculations, rankings, benchmarking, and comparisons. These are required features to build better Information Systems for sustainability. The use of XBRL by Information Systems represents progress towards the improvement of internal processes at strategic, management and operational level inside companies. Financial and non-financial reporting initiatives trust XBRL as the best way to standardise, exchange and validate the information that they manage. Thus, Information Systems using XBRL can respond to stakeholder demands in areas of financial and environmental sustainability, from the step of initial data preparation, subsequent disclosure through to effective data-driven management.

The use of XBRL for non-financial reporting by Information Systems is not happening yet. One reason for this is that XBRL is still in the initial implementation stage for some non-mandatory initiatives and there is a lack of action to move XBRL towards the potential audiences. For example, CDP and GRI have been publishing their XBRL taxonomies to the public since 2012 and 2010 respectively, however, they have not yet evolved their systems to be able to accept, validate or publish information in XBRL or other open data standards. XBRL, like any other data standard, needs the engagement of the open and private software community to develop tools and Information Systems to bring the benefits and value of XBRL to companies and their stakeholders, including investors, governments, suppliers and academics. CDP is currently taking further actions, such as updating its annual taxonomies, evolving its reporting system, and working more closely with the software community and their stakeholders on their use of XBRL as a way to increase the adoption of its data.

7.1.4 Conclusions

The TCFD has a clear mandate from governments, companies and investors to improve the disclosure landscape so that they can better manage climate-related financial risks. However, several challenges have to be addressed to increase the use of climate-related information for the purposes of financial and corporate decision making. Decisions depend on information, and XBRL provides a mechanism to drive and strengthen data-driven decision making that is inclusive of climate and financial aspects, thereby overcoming the pending challenges on climate-related financial reporting. It offers the following benefits:

- 1) It connects the environmental and financial information models, increasing their significance for the purpose of decision making;
- 2) It reduces the costs of reporting, enabling data to be consistent, structured and usable across different existing disclosure regimes;

- 3) It improves data quality, facilitates data validation, allows for comparisons against external data, and provides more transparency in financial and extra-financial reporting; and
- 4) It enhances the usefulness of the information, which supports better information systems to drive short and long-term decisions.

The inclusion of XBRL as a key digital reporting technology within the TCFD recommendations could provide both a potential mechanism to overcome existing challenges and maximise the future development of a sustainable economy by providing a better picture of the financial risks associated with climate change and other sustainability matters, strengthening the dialogue between companies and shareholders and evolving it to include a broader set of stakeholders, including global public opinion.

7.2 Using XBRL for representing the CDP environmental data: insights and modelling decisions

7.2.1 Motivation

The development of the CDP taxonomy has been part of the innovation strategy to better utilise CDP data in business strategies and operations. It represents a starting point for organisations to create value using digital reporting standards and help incorporate Climate Change data into decision making.

The following sections explain (1) how the 2016 Climate Change questionnaire is modelled and its requirements, and (2) the technical implementation in XBRL format.

7.2.2 Data modelling: 2016 climate change questionnaire technical requirements

This section details technical aspects required by the 2016 Climate Change questionnaire, regarding data modelling and business rules. The purpose of this section is to provide a better understanding of the technical basis of CDP's questionnaires for early adopters. The taxonomy

was created using the Fujitsu XWand Taxonomy Editor software²⁷ and the methodology followed was described in section 2.2.

The 2016 Climate Change questionnaire is composed of a set of common questions valid for any sector-related company, and specific questions for the following sectors:

- (1) Electric Utility
- (2) Auto and Auto Component Manufacture
- (3) Oil and Gas
- (4) Information and Communications Technology
- (5) Food, Beverage and Tobacco

The design of the questionnaire implies six different entry points to deal with Climate Change data depending on the sector-related company (Figure 19). As this was a pilot study, only the Common Module and Oil and Gas (sector module) have been implemented.

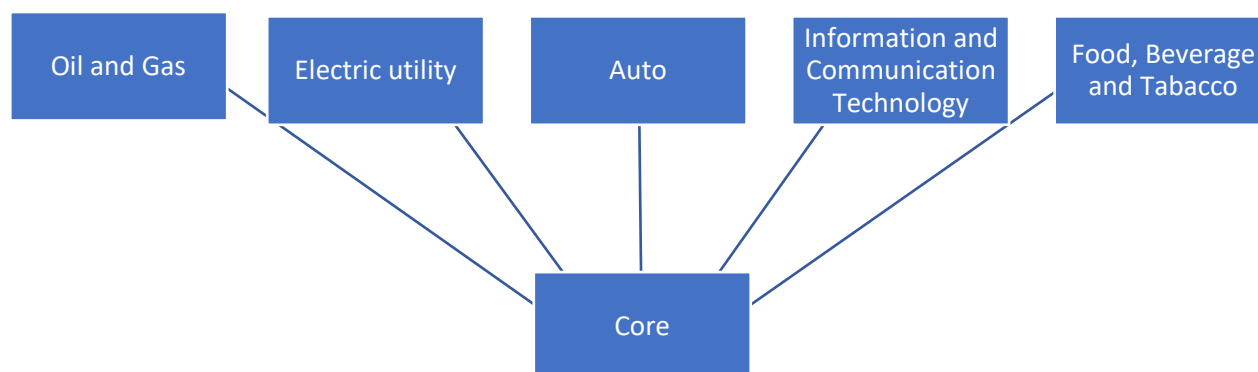


Figure 19. 2016 Climate Change entry points

²⁷ Fujitsu Xwand Taxonomy Editor: <https://www.fujitsu.com/global/products/software/middleware/application-infrastructure/interstage/xbrltools/downloadbiz21.html> [Accessed August 2019]

7.2.2.1 Core Module (related to all sectors)

The core module is composed of a set of modules: General information, Management, Risk and Opportunities and Emissions. Each module consists of a set of sections that are addressed by different questions, which is represented in the following diagram (Figure 20):

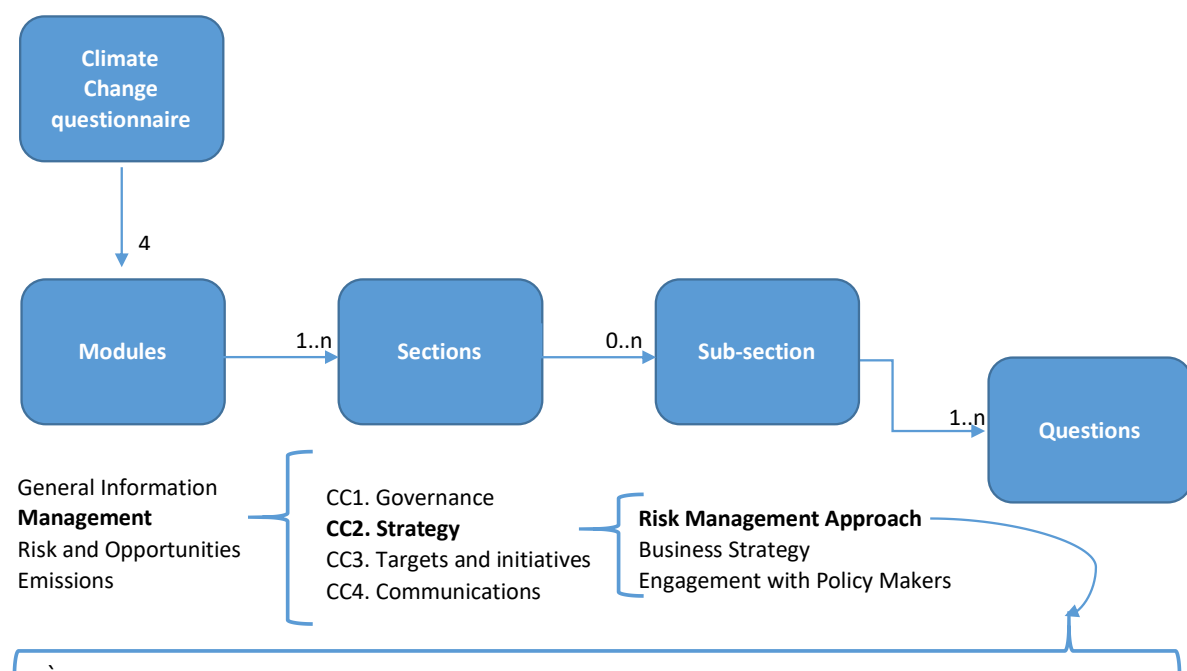


Figure 20. Climate Change questionnaire structure

CC2.1 Please select the option that best describes your risk management procedures with regard to climate change risks and opportunities

CC2.1a Please provide further details on your risk management procedures with regard to climate change risks and opportunities

CC2.1b Please describe how your risk and opportunity identification processes are applied at both company and asset level

CC2.1c How do you prioritize the risks and opportunities identified?

CC2.1d Please explain why you do not have a process in place for assessing and managing risks and opportunities from climate change, and whether you plan to introduce such a process in the future

Table 6 shows the framework of the questionnaire, making its structure and content clearer.

Module	Section	Subsection	Question
General Information	CC0. Introduction page		
Management	CC1. Governance	Group and Individual Responsibility	CC1.1, CC1.1a
		Individual performance	CC1.2, CC1.2a
	CC2. Strategy	Risk Management approach	CC2.1, CC2.1a, CC2.1b, CC2.1c, CC2.1d
		Business Strategy	CC2.2, CC2.2a, CC2.2b, CC2.2c, CC2.2d
		Engagement with Policy Makers	CC2.3, CC2.3a, CC2.3b, CC2.3c, CC2.3d, CC2.3e, CC2.3f, CC2.3g
	CC3. Targets and Initiatives	Targets	CC3.1, CC3.1a, CC3.1b, CC3.1c, CC3.1d, CC3.1e, CC3.1f
		Emissions Reduction Initiatives	CC3.2, CC3.2a, CC3.3, CC3.3a, CC3.3b, CC3.3c, C3.3d
	CC4. Communications	Communications	CC4.1
Risk and Opportunities	CC5. Climate Change Risks	Risk	CC5.1
	CC6. Climate Change Opportunities	Opportunities	CC6.1
Emissions	CC7. Emissions Methodology	Base Year	CC7.1
		Methodology	CC7.2, CC7.2a, CC7.3, CC7.4
	CC8. Emissions data	Boundary	CC8.1
		Scope 1 and 2 Emissions data	CC8.2, CC8.3, CC8.4, CC8.4a
		Data Accuracy	CC8.5

		External verification or assurance	CC8.6, CC8.6a, CC8.6b, CC8.7, CC8.7a, CC8.8
		Carbon Dioxide Emissions from Biologically Sequestered Carbon	CC8.9
	CC9. Scope 1 Emissions Breakdown	Scope 1 Emissions Breakdown	CC9.1, CC9.1a, CC9.2, CC9.2a, CC9.2b, CC9.2c, CC9.2d
	CC10. Scope 2 Emissions Breakdown	Scope 2 Emissions Breakdown	CC10.1, CC10.1a, CC10.2, CC10.2a, CC10.2b, CC10.2c
	CC11. Energy	Energy	CC11.1, CC11.2, CC11.3, CC11.3a, CC11.4, CC11.5
	CC12. Emissions Performance	Emission Performance	CC12.1, CC12.1a, CC12.1b
		Emissions Intensity	CC12.2, CC12.3
	CC13. Emissions Trading	Emissions Trading	CC13.1, CC13.1a, CC13.1b, CC13.2, CC13.2a
	CC14. Scope 3 emissions	Scope 3 emissions	CC14.1, CC14.2, CC14.2a, CC14.3, CC14.3a, CC14.4, CC14.4a, CC14.4b, CC14.4c, CC14.4d

Table 6. Climate Change Questionnaire 2016 (common module skeleton)

7.2.2.2 Sector specific modules – Oil and Gas

More information about the Oil and Gas sector-specific module can be found at the following link: <https://www.cdp.net/Documents/Guidance/2016/CDP-2016-Oil-Gas-Module-Reporting-Guidance.pdf>

Table 7 presents the Oil and Gas framework. This module is an extension of the Common Module.

Section	Question
---------	----------

OG0. Reference Information	OG0.1
OG1. Production and reserves by hydrocarbon types	OG1.1, OG1.2, OG1.3, OG1.4, OG1.5, OG1.6, OG1.6a, OG1.6b,
OG2. Emissions by segment in the O&G value chain	OG2.1, OG2.2, OG2.3, OG2.4
OG3. Scope 1 emissions by emissions category	OG3.1, OG3.2, OG3.3, OG3.4
OG4. Transfers and sequestration of CO2 emissions	OG4.1, OG4.2, OG4.3, OG4.4, OG4.5, OG4.6, OG4.7, OG4.8, OG4.9
OG5. Sales and emissions intensity	OG5.1, OG5.2, OG5.3
OG6. Development strategy	OG6.1, OG6.2, OG6.3
OG7. Methane from the natural gas value chain	OG7.1, OG7.2, OG7.3, OG7.3a, OG7.3b, OG7.4, OG7.5, OG7.6, OG7.6a, OG7.7, OG7.7a, OG7.7b

Table 7. Oil and Gas sector specific module skeleton

7.2.2.3 Disclosure pathway (lead and led disclosures)

The Climate Change questionnaire uses specific functionality to guide preparers of information through its disclosure. We call this functionality the “disclosure pathway”, as it indicates the path that a preparer should follow during the disclosure process. The advantages of this approach are:

- It indicates the order and dependencies between data points;
- It guides the preparer during the disclosure process;
- It allows construction of data structures, transcending from high-level to the detailed-level;
- It speeds up the evaluation of the final report.

Figure 21 illustrates the questionnaire’s pathway:

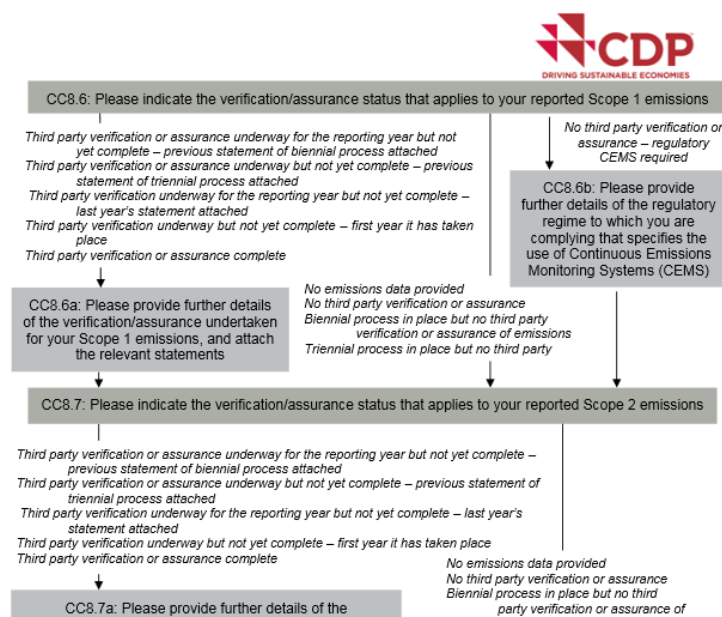


Figure 21. Question pathway

For example, to answer the question 8.6a or 8.6b, question 8.6 must be answered first: “whether the GHG emissions have been verified (yes or no)”. This type of disclosure is called a “lead disclosure” due to the influence it has on subsequent questions to be answered, i.e. 8.6a or 8.6b, which is called “led disclosure”. A “yes” response to “lead disclosure” 8.6 will lead to “led disclosure” 8.6a and “no” will lead to “led disclosure” 8.6b.

7.2.2.4 Enumeration list/list of values

The Climate Change questionnaire uses the “list of values” structure in order to facilitate the analysis of textual data. It is important to note that these structures are normally reused during the course of the questionnaire.

In Figure 22, we present an enumeration structure illustrated by question OG3.1, where the element **Segment**, is a list composed of five values: (1) exploration, production and gas processing; (2) storage, transportation and distribution; (3) speciality operations; (4) refining; (5) retail and marketing.

OG3.1 Please confirm the consolidation basis (financial control, operational control, equity share) used to report Scope 1 emissions by emissions category

Segment	Consolidation basis for reporting Scope 1 emissions by emissions category
Please select...	Please select...
Exploration, production & gas processing	
Storage, transportation & distribution	
Specialty operations	
Refining	
Retail & marketing	

OG3.2 If different consolidation bases have been used to report by emissions categories (combustion, flaring, process emissions, fugitive emissions) in the various segments

Add Row(s) 1

Figure 22. Example of enumeration structure in CDP questionnaire

7.2.2.5 Validation rules

The Climate Change questionnaire requires validations rules in order to ensure completeness and data consistency. It requires two different type of validations:

1. Some data fields must be reported with appropriate information and cannot be left blank, such as question CC1.1.
2. Some data fields must comply with arithmetical relationships. For example, the sum of the values provided for the fuel breakdown (question CC11.3 column 2) must be equal to the sum of values provided for all energy types (question CC11.2 column 2).

7.2.2.6 Units

The Climate Change questionnaire requires data in different units:

- **Energy type:** MWh, GWh, BOE
- **Mass type:** Metric tonnes CO2e
- **Monetary type:** USD, EUR
- **Power type:** MW
- **Area type:** m²

Data input by the user in these fields must comply with the stipulated units.

7.2.2.7 Additional information: Attached external document

Some questions require additional information in the form of attached documents such as the question CC7.4 presented in Figure 23.

CC7.4 Please give the emissions factors you have applied and their origin; alternatively, please attach an Excel spreadsheet with this data at the bottom of this page (CDP 2013 Q7.4, amended)

Fuel/Material/Energy	Emission Factor	Unit	Reference

Figure 23. Example of attached document structure

These types of questions try to facilitate the traceability and authenticity of the information provided, which are useful for verification and assurance purposes.

7.2.3 Modelling CDP concepts in XBRL

7.2.3.1 Summary of the taxonomy

The Climate Change Taxonomy contains a dictionary with 2030 elements distributed as follows:

Common module:

Reporting items: 222

Presentation items: 322

Dimensional structures: 87

List of values/ Enumerations: 848

List of regions: 50

List of countries: 235

Oil and Gas module:

Reporting items: 75

Presentation items: 65

Dimensional structures: 36

List of values/ Enumerations: 90

This number of elements is sufficient to cover a complete Climate Change Report and Oil and Gas module, according to 2016 Climate Change questionnaire.

7.2.3.2 Data dictionary model

For each element of the data dictionary, the following definitions can be found:

- **Semantic name of the concept.** This name is composed of characters and will be the denomination, in the form of a label, which will be employed in the XBRL reports.

The name is assigned a unique identifier called a 'namespace'. Different concepts cannot have identical names.

- **Metadata identifier,** which is defined uniquely, both for the data dictionary and for the Discoverable Taxonomy Set (DTS). For this, a chain is employed that includes the prefix employed for the namespace.

- **Data type.** This defines the type of data that the label will include. It is possible to choose a data type defined in the type-definition of the XBRL specification or one from the field defined by the XML Schema.

- **State of aggregation.** This characteristic is used to define whether the element is considered on its own (an item) or together with other elements (a tuple).

- **State of definition.** This indicates if the element has all the characteristics and types completely defined. If it is not complete, it is defined as an abstract element.

- **Form of measurement.** The value included in a label of the Climate Change report in XBRL can be considered as the measurement made at one moment in time or for a specified period.

7.2.3.2.1 Data model for the presentation linkbase

The presentation linkbase contains the relationships existing between the elements within the data dictionary, with the purpose of enabling the elements grouped together by extended links

to be represented hierarchically. Table 8 and 9 present the correspondences between the presentation linkbase and the Core and Oil and Gas framework.

A. Core module

General information	
Presentation Linkbase	CDP information request
[00000000] General information about the report	Section CC0. Introduction page Sign-off :CC15.1 *Note:

Management	
Presentation Linkbase	CDP information request
[01000000] Disclosure of governance	Section CC1. Governance Group and Individual Responsibility: CC1.1, CC1.1a Individual performance: CC1.2, CC1.2a
[02000000] Disclosure of strategy	Section CC2. Strategy Risk Management Approach: CC2.1, CC2.1a, CC2.1b, C2.1c, cc2.1d Business Strategy: CC2.2, CC2.2a, CC2.2b Engagement with Policy Makers: CC2.3, CC2.3a, CC2.3b, CC2.3c, CC2.3d, CC2.3e, CC2.3f, CC2.3g, CC2.3h, CC2.3i
[03000000] Disclosure of targets and initiatives	Section CC3. Targets and Initiatives Targets: CC3.1, CC3.1a, CC3.1b, CC3.1c, CC3.1d, CC3.1e, Emission Reduction Initiatives: CC3.2, CC3.2a, CC3.3, CC3.3a, CC3.3b, CC3.3c, CC3.3d
[04000000] Disclosure of communication	Section CC4. Communication Communication: CC4.1

Risks and Opportunities

Presentation Linkbase	CDP information request
[0500000] Disclosure of Climate Change risks	Section CC5. Climate Change risks Climate Change risks: CC5.1
[0600000] Disclosure of Climate Change opportunities	Section CC6. Climate Change opportunities Climate Change opportunities: CC6.1

Emissions	
Presentation Linkbase	CDP information request
[0700000] Disclosure of emission methodology	Section CC7. Emissions Methodology Base year: CC7.1 Methodology: CC7.2, CC7.2a, CC7.3, CC7.4
[0800000] Disclosure of emission data	Section CC8. Emissions Data Boundary: CC8.1 Scope 1 and 2 Emissions data: CC8.2, CC8.3, CC8.4, CC8.4a Data accuracy: CC8.5, External Verification or Assurance: CC8.6, CC8.6a, CC8.6b, CC8.7, CC8.7a, CC8.8 Carbon Dioxide Emissions from Biologically Sequestered Carbon: CC8.9, CC8.9a
[0900000] [1000000] Disclosure of scope 1 and Scope 2 emissions breakdowns	Section CC9. Scope 1 Emissions Breakdown Scope 1 Emissions Breakdown: CC9.1, CC9.1a, CC9.2, CC9.2a, CC9.2b, CC9.2c, CC9.2d, CC9.2e, Section CC10. Scope 2 Emissions Breakdown Scope 2 Emissions Breakdown: CC10.1, CC10.1a, CC10.2, CC10.2a, CC10.2b, CC10.2c, CC10.2d
[1100000] Disclosure of energy	Section CC11. Energy Energy: CC11.1, CC11.2, CC11.3, CC11.4
[1200000] Disclosure of emissions performance	Section CC12. Emissions Performance Emissions History: CC12.1, CC12.1a Emissions Intensity: CC12.2, CC12.3, CC12.4
[1300000] Disclosure of emissions trading	Section CC13. Emissions Trading Emissions Trading: CC13.1, CC13.1a, CC13.1b, CC13.2, CC13.2a

[14000000] Disclosure of Scope 3 emissions	Section CC14. Scope 3 Emissions Scope 3 emissions: CC14.1, CC14.2, cc14.2a, CC14.3, CC14.3a, CC14.4, CC14.4a, CC14.4b, CC14.4c, CC14.4d
--	--

Table 8. Correspondence between presentation linkbase and Core module framework

B. Oil and Gas module

Oil and Gas module	
Presentation Linkbase	CDP information request
[OG00000000] Disclosure of reference information	Section OG0: Reference information OG0.1
[OG01000000] Disclosure of production and reserves of hydrocarbon	Section OG1: Production & reserves by hydrocarbon type OG1.1, OG1.2, OG1.3, OG1.4, OG1.5, OG1.6, OG1.6a, OG1.6b,
[OG02000000] Disclosure of emissions by segment in the Oil and Gas value chain	Section OG2: Emissions by segment in the O&G value chain OG2.1, OG2.2, OG2.3, OG2.4
[OG03000000] Disclosure of scope 1 emissions category	Section OG3: Scope 1 emissions by emissions category OG3.1, OG3.2, OG3.3, OG3.4
[OG04000000] Disclosure of transfers and sequestrations of CO2 emissions	Section OG4: Transfers & sequestration of CO2 emissions OG4.1, OG4.2, OG4.3, OG4.4, OG4.5, OG4.6, OG4.7, OG4.8, OG4.9
[OG05000000] Disclosure of sales and emissions intensity	Section OG5: Sales and emissions intensity OG5.1, OG5.2, OG5.3
[OG06000000] Disclosure of development strategy	Section OG6: Development strategy OG6.1, OG6.2, OG6.3
[OG07000000] Disclosure of methane from the natural gas value chain	Section OG7: Methane from the natural gas value chain OG7.1, OG7.2, OG7.3, OG7.3a, OG7.3b, OG7.4, OG7.5, OG7.6, OG7.6a, OG7.7, OG7.7a, OG7.7b

Table 9. Correspondences between presentation linkbase and Oil and Gas module framework

7.2.3.2.2 Data model for the definition linkbase

The definition linkbase contains the relationships between the concepts and dimensional structures to define each table.

A. Common module

General information	
Presentation Linkbase	Definition Linkbase
[00000000] General information about the report	[00100000] Disclosure of person responsible for signing off the report

Management	
Presentation Linkbase	Definition Linkbase
[01000000] Disclosure of governance	[01200000a] Disclosure of incentives
[02000000] Disclosure of strategy	[02300000a] Disclosure of direct activities policy makers [02300000c] Disclosure of trade associations with position on climate change legislation
[03000000] Disclosure of targets and initiatives	[03100000a] Disclosure of absolute targets [03100000b] Disclosure of intensity targets [03100000c] Disclosure of renewable energy consumption production target [03200000a] Disclosure of products service low carbon [03300000a] Disclosure of emission reduction projects and estimated reduction [03300000b] Disclosure of emission reduction initiatives in reporting period [03300000c] Disclosure of methods used to drive investments in emission reduction activities
[04000000] Disclosure of communication	[04100000] Disclosure of type of publications

Risks and Opportunities	
Presentation Linkbase	Definition Linkbase
[0500000] Disclosure of Climate Change risks	[05100000abc] Disclosure of information on climate change risks
[0600000] Disclosure of Climate Change opportunities	[06100000abc] Disclosure of information on climate change opportunities

Emissions	
Presentation Linkbase	Definition Linkbase
[0700000] Disclosure of emission methodology	[07100000] Disclosure of base year emissions [07200000] Disclosure of standards, protocols or methodologies used to collect activity data and calculate scope 1 and 2 emissions [07300000] Disclosure of global warming potentials used [07400000] Disclosure of emission factors used
[0800000] Disclosure of emission data	[08200000] Disclosure of total gross scope 1 and scope 2 emission data [08400000a] Disclosure of greenhouse gas inventory exclusions [08500000] Disclosure of uncertainty scope 1 and scope 2 emission data [08600000a] Disclosure of detailed information on verification or assurance applicable to scope 1 and scope 2 emissions [08600000b] Disclosure of the regulatory regime specifying the use of Continuous Emissions Monitoring Systems [08800000] Disclosure of information on verification assurance of additional data points other than emissions figures
[09000000][10000000] Disclosure of scope 1 and scope 2 emissions breakdowns	[09100000a] [10100000a] Disclosure of scope 1 and scope 2 emissions breakdown by country and region [09200000a][10200000a] Disclosure of scope 1 and scope 2 emissions breakdown by business division

	<p>[09200000b][10200000b] Disclosure of scope 1 and scope 2 emissions breakdown by facility</p> <p>[09200000c] Disclosure of scope 1 emissions breakdown by greenhouse gas</p> <p>[09200000d][10200000c] Disclosure of scope 1 and scope 2 emissions breakdown by activity</p> <p>[10100000a] Disclosure of energy consumption associated with scope 2 emissions breakdown by country and region</p>
[1100000] Disclosure of energy	<p>[11200000] Disclosure of energy</p> <p>[11300000] Disclosure of fuel consumption</p> <p>[11400000] Disclosures on accounting practices of low carbon electricity, heat, steam or cooling in scope 2.</p>
[1200000] Disclosure of emissions performance	<p>[12100000a] Disclosure of reasons for change in gross global emissions</p> <p>[12200000][12300000] Disclosure of emissions intensity</p>
[1300000] Disclosure of emissions trading	<p>[13100000a] Disclosure of information on participation in emissions trading schemes</p> <p>[13200000a] Disclosure of use of project-based carbon credits within reporting period</p>
[1400000] Disclosure of Scope 3 emissions	<p>[14100000] Disclosure of scope 3 emissions by category</p> <p>[14200000a] Disclosure of scope 3 verification</p> <p>[14300000a] Disclosure of changes in absolute scope 3 emissions compared to previous reporting period</p> <p>[14400000c] Disclosure of how GHG emissions and climate change strategy data from suppliers is used</p>

Table 10. Definition linkbase – Common module

This linkbase also contains the definition of the list of values structures using Extensible Enumerations; see section 6.2.3.5 for more detail about the use of Extensible enumerations.

List of values structures - Extensible Enumerations

Definition Linkbase
[99000001] Type of beneficiary [Enumeration]
[99000002] Type of incentives [Enumeration]
[99000003] Type of highest level of responsibility [Enumeration]
[99000004] Type of timeframe [Enumeration]
[99000006] Type of engagement activities legislation [Enumeration]
[99000007] Type of engagement activities policy makers [Enumeration]
[99000008] Type of consistency position between entity and trade association [Enumeration]
[99000009] Type of frequency risk and opportunities [Enumeration]
[99000005] Type of addressee reports [Enumeration]
[99000011] Type of metrics [Enumeration]
[99000012] Type of direction [Enumeration]
[99000013] Type of emissions reduction activities [Enumeration]
[99000014] Type of payback period [Enumeration]
[99000015] Type of method used to drive investment in reduction [Enumeration]
[99000016] Type of publications [Enumeration]
[99000018] Type of potential impact [Enumeration]
[99000019] Type of direct / indirect [Enumeration]
[99000020] Type of likelihood [Enumeration]
[99000021] Type of magnitude impact [Enumeration]
[99000058] Type of risks [enumeration]
[99000057] Type of risks drivers [enumeration]
[99000056] Type of opportunities drivers [enumeration]
[99000059] Type of opportunities [enumeration]
[99000060] Type of potential impacts [enumeration]
[99000055] Type of timeframe [enumeration]
[99000022] Type of standard, protocol or methodology used to collect activity data and calculate scope 1 and 2 emissions [enumeration]
[99000023] Type of Greenhouse gas [enumerations]
[99000024] Type of Greenhouse gases global warming potentials references [enumeration]
[99000025] Type of materials, fuels and electricity [enumeration]
[99000026] Type of unit, numerator [enumeration]
[99000027] Type of unit, denominator [enumeration]
[99000028] Type of materials, fuels and electricity [enumeration]
[99000029] Type of greenhouse gas inventory boundary [enumeration]
[99000030] Type of uncertainty range [enumeration]
[99000032] Type of verification or assurance status applicable to scope 1 emissions [enumeration]
[99000033] Type of verification or assurance [enumeration]
[99000034] Type of relevant standard [enumeration]
[99000035] Type of regulation [enumeration]
[99000036] Type of additional data points verified [enumeration]
[99000037] Country [enumeration]

[99000038] Region [enumeration]
[99000040] Type of fuel [enumeration]
[99000041] Type of basis for applying a low carbon emission factor [Enumeration]
[99000042] Type of energy [enumeration]
[99000043] Type of reason for change [enumeration]
[99000045] Type of evolution [enumeration]
[99000046] Type of emissions trading schemes [enumeration]
[99000047] Type of ownership [enumeration]
[99000048] Type of project carbon credit
[99000049] Type of verified to standard carbon credit
[99000050] Type of purpose carbon credit
[99000051] Type of scope 3 emissions categories [enumeration]
[99000054] Type of use of GHG emissions and climate change strategy data from suppliers [enumeration]
[99000052] Type of engagements with elements of value chain on GHG emissions and climate change strategies [enumeration]
[99000053] Type of ability to compare Scopes emissions for the reporting year with those of previous year [enumeration]
[99000055] Type of methods used to drive investment in emission reduction activities [enumeration]
[99000017] Type of risk management procedures related to climate change risks and opportunities [enumeration]
[99000062] Relevance of exclusions [enumeration]
[99000063] Scope 3 Evaluation Status [enumeration]
[99000064] Type of Incentivized Performance Indicator [enumeration]
[99000065] Type of reason for not having a process in place for assessing and managing risks and opportunities from climate change [enumeration]
[99000066] Science target enumeration
[99000067] Use internal price carbon [enumeration]
[99000069] Type of energy [enumeration]
[99000070] Type of responses [enumeration]
[99000080] Type of aggregations [enumeration]
[99000081] Type of responses low carbon [enumeration]
[99000082] Type of projects, methodologies low carbon [enumeration]
[99000083] Type of R&D carbon products [enumeration]
[99000084] Type of publication status [enumeration]
[99000085] Type of scopes emissions methodology [enumeration]
[99000086] Type of cycles [enumeration]
[99000087] Type of status verification [enumeration]
[99000088] Type of scope 2 emissions to base emissions performance calculations [enumeration]
[99000089] Type of source of scope 3 emissions [enumeration]
[99000090] Lifetime initiative [enumeration]

[99000091] Stage of development [enumeration]
 [99000092] Operational spend on energy [enumeration]
 [99000093] Carbon credit origination or carbon purchase [enumeration]
 [99900000] Adimensional
 [99000094] Type incentivized performance indicator [enumeration]
 [99000095] Type scope emissions [enumeration]
 [99000096] Type disclosure [enumeration]
 [99000097] Type source uncertainty [enumeration]

Table 11. Extensible enumerations within the definition linkbase

B. Oil and Gas module

Oil and Gas module	
Presentation Linkbase	Definition Linkbase
[OG00000000] Disclosure of reference information	
[OG01000000] Disclosure of production and reserves of hydrocarbon	[OG01200000] Disclosure of gross and net production of hydrocarbon [OG01300000] Disclosure of reserves of hydrocarbon [OG01500000] Disclosure of breakeven cost hydrocarbon reserves
[OG02000000] Disclosure of emissions by segment in the Oil and Gas value chain	[OG02100000] Disclosure of consolidation basis used to report the Scope 1 and Scope 2 by segment in the O&G value chain [OG02300000] Disclosure of masses of gross Scope 1 broken down by value chain segment [OG02400000] Disclosure of masses of gross Scope 2 broken down by value chain segment
[OG03000000] Disclosure of scope 1 emissions category	[OG03100000] Disclosure of consolidation basis used to report the Scope 1 and Scope 2 by category [OG03300000] Disclosure of masses of gross Scope 1 released into the atmosphere for the whole organisation broken down by emissions category
[OG04000000] Disclosure of transfers and sequestrations of CO2 emissions	[OG04700000] Disclosure of gross CO2 captured for purposes of carbon capture and sequestration

	[OG04800000] Disclosure of masses CO2 injected and stored for purposes of CCS
[OG05000000] Disclosure of sales and emissions intensity	[OG05100000] Disclosure of hydrocarbon types [OG05200000] Disclosure of estimated emissions intensities (Scope 1 + Scope 2)
[OG06000000] Disclosure of development strategy	[OG06100000] Disclosure of strategic development area providing financial information [OG06200000] Disclosure of future capital expenditure plans for different strategic development areas [OG06300000] Disclosure of current expenses in Research and Development and future expenses
[OG07000000] Disclosure of methane from the natural gas value chain	[OG07100000] Disclosure of consolidation basis [OG07400000] Disclosure of methane emissions inventory estimated [OG07500000] Disclosure of methane emissions rate

List of values structures - Extensible Enumerations

Definition Linkbase

[OG09900001] Type of petroleum industry [enumeration]
 [OG09900002] Type of hydrocarbon [enumeration]
 [OG09900003] Type of consolidation boundary [enumeration]
 [OG09900004] Type of status [enumeration]
 [OG09900005] Type of segment [enumeration]
 [OG09900006] Type of emissions category [enumeration]
 [OG09900007] Type of capture pathway in CCS [enumeration]
 [OG09900008] Type of injection and storage pathway [enumeration]
 [OG09900009] Type of strategic development area [enumeration]
 [OG09900010] Type of proportion [enumeration]
 [OG09900011] Type of areas [enumeration]
 [OG09900012] Type of methane emissions methodology [enumeration]

Table 12. Definition linkbase - Oil and Gas module

7.2.3.2.3 Data model for the Reference Linkbase

The Reference Linkbase contains the relationships between the elements of the data dictionary and the documentary references.

Reference part	Use in the taxonomy
Publisher	Publisher of the reference material, such as {CDP}
Paragraph	Question unique identifier, e.g. CC0.1
Note	Question description, e.g. Provide a general introduction to your organisation and operations, summarizing the GHG emissions inventory.
URI	Full URI of the reference document, e.g. https://www.cdp.net/Documents/Guidance/2016/CDP-2016-Climate-Change-Reporting-Guidance.pdf

Table 13. Reference linkbase

7.2.3.2.4 Data model for the Formula Linkbase

The Formula Linkbase contains the relationships to define mathematical rules for validation.

The use of formulas in this taxonomy provides sophisticated validations constraints, with a full set of mathematical functions.

7.2.3.3 Modularity

As illustrated in Figure 24 below, the Climate Change taxonomy is organised logically. This version contains two schema files to serve as entry points: the Common Module (cdp-cc-2016-08-30.xsd) and Oil and Gas Sector Specific Module (cdp-og-2016-08-30.xsd).

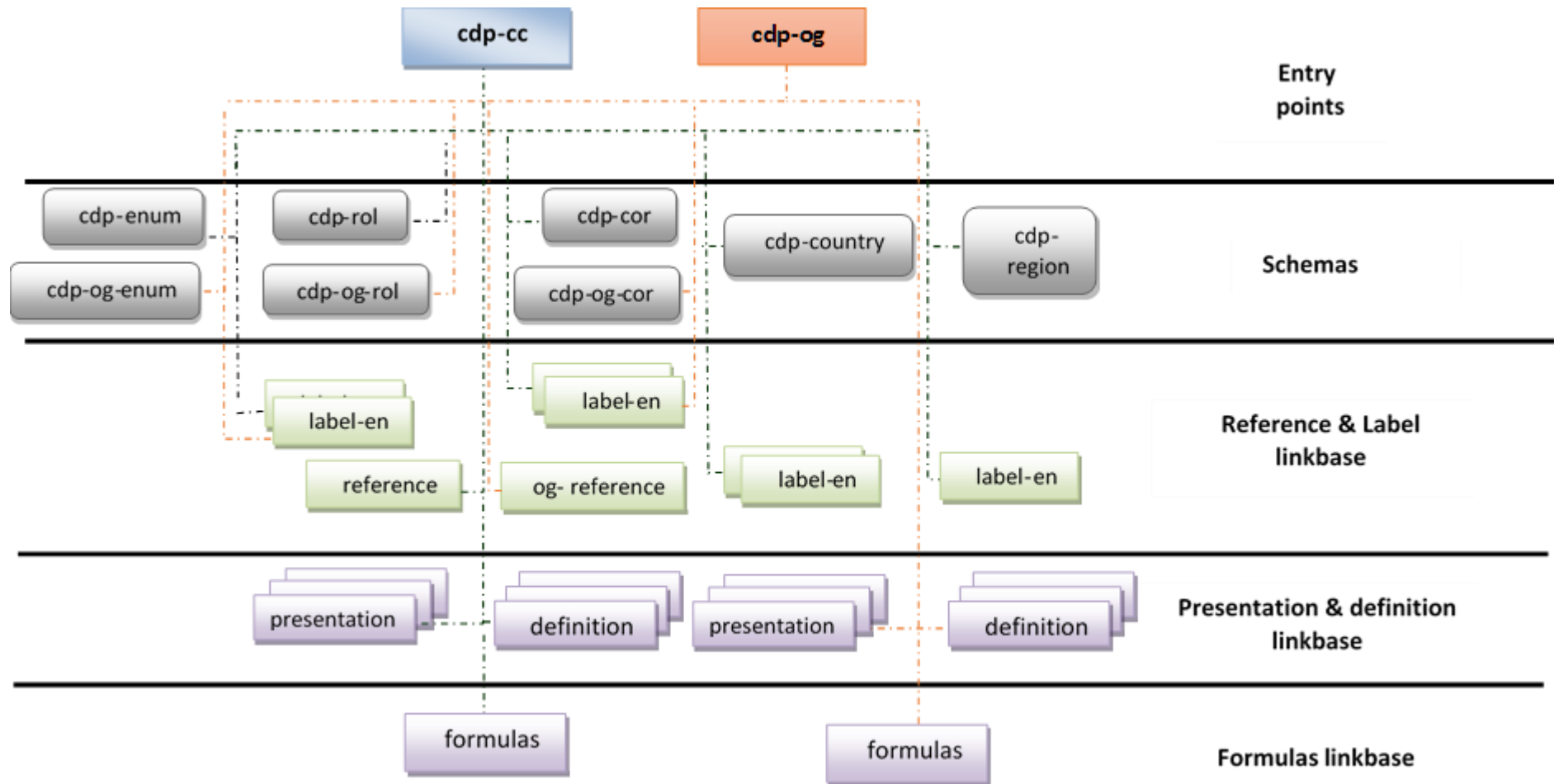


Figure 24. Climate Change Taxonomy physical architecture

7.2.3.4 Dimensions

Dimensions 1.0²⁸ enriches rules and procedures for constructing dimensional taxonomies and therefore instance documents. Taxonomies using the Dimensions specification can define new dimensional contexts, specifying valid values ("domains") for dimensions, using a mechanism called a hypercube to define which dimensions apply to which business concept. There are two types of dimensions:

- Explicit dimensions: These have a fixed number of dimension members. For example, in a two-dimensional table, the number of row and columns are known.
- Typed dimensions: the number of dimension members is unknown. For example, in a two-dimensional table, it means that the number of columns is known, but the number of rows depends on user reporting needs.

In this taxonomy, typed dimensions are the only ones implemented.

7.2.3.5 Extensible Enumerations

Extensible Enumerations allow the definition of the list of values for primary reporting concepts. It allows:

- The creation of enumeration values in each language;
- The creation of a list of value structures that can be reused;
- Disclose facts with multi-valued enumerations (*new in Extensible Enumeration 1.1*).

This taxonomy uses extensible enumerations 1.1²⁹ (*proposed recommendation 8th February 2017*) to include and reuse the list of values concepts in dimensional structures.

The following is an example of a list of values for "Highest level of responsibility for climate change within entity":

- Individual / Sub-set of the Board or other committee appointed by the Board

²⁸ <http://www.xbrl.org/specification/dimensions/rec-2012-01-25/dimensions-rec-2006-09-18+corrected-errata-2012-01-25-clean.html> [Accessed August 2019]

²⁹ <https://www.xbrl.org/Specification/extensible-enumerations-1.1/PR-2017-02-08/extensible-enumerations-1.1-PR-2017-02-08.html> [Accessed August 2019]

- Senior Manager / Officer
- Other Manager / Officer
- No individual or committee with overall responsibility for climate change

Example of an enumeration concept (type=EnumerationItemType) – (cdp-cor-2016-08-30.xsd):

```
<xsd:element
name="HighestLevelDirectResponsabilityClimateChangeEntityLeadDisclosu
re"
id="cdp_HighestLevelDirectResponsabilityClimateChangeEntityLeadDisclo
sure" type="enum:enumerationItemType"
enum:domain="cdp:HighestLevelDirectResponsabilityClimateChangeEntityE
numerationList" substitutionGroup="xbrli:item"
enum:linkrole="http://cdp-xbrl.net/2016-08-
30/ELR/TypeHighestLevelResponsability " nillable="true"
xbrli:periodType="duration"/>
  <xsd:element
name="HighestLevelDirectResponsabilityClimateChangeEntityEnumerationL
ist"
id="cdp_HighestLevelDirectResponsabilityClimateChangeEntityEnumeratio
nList" type="xbrli:stringItemType" substitutionGroup="xbrli:item"
abstract="true" nillable="true" xbrli:periodType="duration"/>
```

Definition of the list of values – (cdp-def-9900003-2016-08-30.xml):

```
<link:definitionLink xlink:type="extended" xlink:role="http://cdp-
xbrl.net/2016-08-30/ELR/TypeHighestLevelResponsability">
  <link:loc xlink:type="locator" xlink:href="../../cor/cdp-cor-2016-08-
30.xsd#cdp_HighestLevelDirectResponsabilityClimateChangeEntityEnumerationLis
t"
  xlink:label="HighestLevelDirectResponsabilityClimateChangeEntityEnumerationL
ist"
  xlink:title="HighestLevelDirectResponsabilityClimateChangeEntityEnumerationL
ist"/>
  <link:loc xlink:type="locator" xlink:href="../../enumerations/cdp-enum-
2016-08-30.xsd#cdp-enum_NoIndividualCommitteeResponsabilityClimateChange"
  xlink:label="NoIndividualCommitteeResponsabilityClimateChange"
  xlink:title="NoIndividualCommitteeResponsabilityClimateChange"/>
  <link:definitionArc xlink:type="arc"
  xlink:arcrole="http://xbrl.org/int/dim/arcrole/domain-member"
  xlink:from="HighestLevelDirectResponsabilityClimateChangeEntityEnumerationLi
st" xlink:to="NoIndividualCommitteeResponsabilityClimateChange"
  xlink:title="definition:
HighestLevelDirectResponsabilityClimateChangeEntityEnumerationList to
NoIndividualCommitteeResponsabilityClimateChange" order="4.0"/>
  <link:loc xlink:type="locator" xlink:href="../../enumerations/cdp-enum-
2016-08-30.xsd#cdp-enum_OtherManagerOfficer"
  xlink:label="OtherManagerOfficer" xlink:title="OtherManagerOfficer"/>
  <link:definitionArc xlink:type="arc"
  xlink:arcrole="http://xbrl.org/int/dim/arcrole/domain-member"
  xlink:from="HighestLevelDirectResponsabilityClimateChangeEntityEnumerationLi
st" xlink:to="OtherManagerOfficer" xlink:title="definition:
HighestLevelDirectResponsabilityClimateChangeEntityEnumerationList to
OtherManagerOfficer" order="3.0"/>
"
```

Figure 25. Enumeration concept - list of values definition

7.2.3.6 Formulas

Formulas 1.0³⁰ are designed to define validation rules. This taxonomy uses formulas in order to check that the content of the report is valid.

This taxonomy uses two types of rules validations: Existence and Value Assertion.

Existence rules are used to control the completeness of the report. These rules evaluate whether specific data are completed or not. For example:

Disclosure of risk management procedure related to climate change risks and opportunities.
“[MUST] This information must be reported”

³⁰ <http://www.xbrl.org/wgn/xbrl-formula-overview/pwd-2011-12-21/xbrl-formula-overview-wgn-pwd-2011-12-21.html>

```

<variable:variableFilterArc xlink:type="arc"
xlink:arcrole="http://xbrl.org/arcrole/2008/variable-filter"
xlink:from="factVariable_23" xlink:to="conceptName_20" xlink:title="user-
defined: factVariable to conceptName" order="1.0" complement="false"
cover="true"/>
<msg:message xlink:type="resource" xlink:label="message_7"
xlink:role="http://www.xbrl.org/2010/role/message" xlink:title="message"
xml:lang="en" id="label_27">The value of Disclosure of risk management
procedures related to climate change risks and opportunities
[enumeration][lead disclosure] to be completed.</msg:message>
<gen:arc xlink:type="arc"
xlink:arcrole="http://xbrl.org/arcrole/2010/assertion-unsatisfied-message"
xlink:from="existenceAssertion_11" xlink:to="message_7" xlink:title="user-
defined: existenceAssertion to message" order="1.0"/>
<ea:existenceAssertion xlink:type="resource"
xlink:label="existenceAssertion_12" xlink:title="existenceAssertion"
id="existenceAssertion12" aspectModel="dimensional" implicitFiltering="true"
test=". > 0"/>
<variable:factVariable xlink:type="resource" xlink:label="factVariable_24"
xlink:title="factVariable" id="factVariable_21" bindAsSequence="false"/>
<variable:variableArc xlink:type="arc"
xlink:arcrole="http://xbrl.org/arcrole/2008/variable-set"
xlink:from="existenceAssertion_12" xlink:to="factVariable_24"
xlink:title="user-defined: existenceAssertion to factVariable" order="1.0"
name="DisclosureRiskManagementProceduresRelatedClimateChangeRisksOpportunities
LeadDisclosure"/>
<cf:conceptName xlink:type="resource" xlink:label="conceptName_21"
xlink:title="conceptName" id="conceptName_14">
<cf:concept>

<cf:qname>cdp:DisclosureRiskManagementProceduresRelatedClimateChangeRisksOppor
tunitiesLeadDisclosure</cf:qname>
</cf:concept>

```

Figure 26. Formulas - example of existence rule definition

Value Assertion rules are used to ensure the consistency of the report. These rules check expressions, which might involve arithmetic ratios and algorithms. For example:

Scope 1 emissions = \sum (Scope 1 emission per region/country)

Scope 2 emissions = \sum (Scope 2 emission per region/country)

```

<va:valueAssertion xlink:type="resource"
xlink:label="valueAssertion_3" xlink:title="valueAssertion"
id="valueAssertion_01" aspectModel="dimensional"
implicitFiltering="true" test="abs( $EmissionValueGrossCO2e -sum(
$GrossValueByCountryRegionCO2e )) <= $threshold "/>
  <gen:arc xlink:type="arc"
xlink:arcrole="http://xbrl.org/arcrole/2008/assertion-set"
xlink:from="assertionSet" xlink:to="valueAssertion_3"
xlink:title="user-defined: assertionSet to valueAssertion"
priority="0" order="2.0"/>
  <msg:message xlink:type="resource" xlink:label="message_39"
xlink:role="http://www.xbrl.org/2010/role/message"
xlink:title="message" xml:lang="en" id="label">The total emission
value per scope must be equal to the sum of the emissions value
broken down by country/region </msg:message>
  <gen:arc xlink:type="arc"
xlink:arcrole="http://xbrl.org/arcrole/2010/assertion-unsatisfied-
message" xlink:from="valueAssertion_3" xlink:to="message_39"
xlink:title="user-defined: valueAssertion to message" priority="0"
order="1.0"/>
  <variable:factVariable xlink:type="resource"
xlink:label="factVariable_11" xlink:title="factVariable"
id="ScopeTotal" bindAsSequence="false"/>
  <variable:variableArc xlink:type="arc"
xlink:arcrole="http://xbrl.org/arcrole/2008/variable-set"
xlink:from="valueAssertion_3" xlink:to="factVariable_11"
xlink:title="user-defined: valueAssertion to factVariable"
priority="0" order="2.0" name="ScopeTotal"/>
  <cf:conceptName xlink:type="resource" xlink:label="conceptName_7"
xlink:title="conceptName" id="ScopeConcept">
    <cf:concept>
      <cf:qname>cdp:Scope</cf:qname>
    </cf:concept>
  </cf:conceptName>
  <variable:variableFilterArc xlink:type="arc"
xlink:arcrole="http://xbrl.org/arcrole/2008/variable-filter"
xlink:from="factVariable_11" xlink:to="conceptName_7"
xlink:title="user-defined: factVariable to conceptName" priority="0"
order="1.0" complement="false" cover="true"/>
  <df:typedDimension xlink:type="resource"
xlink:label="typedDimension_4" xlink:title="typedDimension"
id="TotalGrossEmissionstTypedDimension">
    <df:dimension>
      <df:qname>cdp:TotalEmissionDataAxis</df:qname>
    </df:dimension>
  </df:typedDimension>
  <variable:variableFilterArc xlink:type="arc"
xlink:arcrole="http://xbrl.org/arcrole/2008/variable-filter"
xlink:from="factVariable_11" xlink:to="typedDimension_4"
xlink:title="user-defined: factVariable to typedDimension"
priority="0" order="2.0" complement="false" cover="true"/>
  <asf:aspectCover xlink:type="resource" xlink:label="aspectCover"
xlink:title="aspectCover" id="IgnoreUnitAspect">
    <asf:aspect>unit</asf:aspect>
  </asf:aspectCover>

```

Figure 27. Formulas- example of value assertion definition

7.2.3.7 Data Types and Units

This taxonomy uses the following data types and units. Instances documents require importing the Unit Registry schema - <http://www.xbrl.org/utr/utr.xml>.

Type Local Name (Taxonomy)	Units (Instance)
stringItemType	
monetaryItemType	UTR schema covers monetary units needed. CDP admits all currencies according to UTR - USD, EUR ,GBP, etc
textBlockItemType	
enumerationItemType	
enumerationsItemType	
booleanItemType	
percentItemType	Percentages shall be reported between 0 and 1, with decimals="4" . A ratio of 18,78% shall be reported as 0.1878 (1878 Basic Points).
gYearItemType	
decimalItemType	TWO DECIMALS, decimals="2" . All monetaryItemType (or derivatives) figures shall be reported in cents: 1755.89 equals 1755.89 Euro, or 1755 Euro and 89 Cents (or the currency actually used).
intItemType	
massItemType	CO2e – This unit must be created by the reporting company. UTR schema does not cover this unit yet. At this moment this is under review by XBRL International.
gYearMonthItemType	
energyItemType	UTR schema covers the units needed. CDP admits all energy units according to UTR - MWh, GWh, BOE
dateItemType	
integerItemType	
Base64BinaryItemType	

Table 14. Data types and units

7.2.4. Style guide and naming convention

This section presents rules that were followed in the creation of the taxonomy so that it was multi-lingual, internally consistent, high-quality and easy-to-use. The main goal of this section is to facilitate the interpretation of the taxonomy and encourage a better understanding of the style guide and naming conventions followed. This is just for informational purposes and no actions are required for preparers to take.

The objectives of these rules are to provide:

- Labels that are easy to use, and identify and relate to reference materials;
- Labels and names that are unique;
- Consistency and predictability, making it easier to locate concepts;
- Support of future translation efforts to achieve consistency between languages;
- Support of future maintenance and changes to the taxonomy.

The style guide will also document exceptions to the defined rules.

7.2.4.1 Naming convention concepts

To facilitate the interpretation of the taxonomy, this section describes the naming conventions used.

a) Concepts MUST have a clear and understandable name

The name given to a concept MUST be clear. A longer and easier to understand name is preferable to a name that is shorter name but more difficult to understand.

b) Concepts SHOULD be created according to the Label Camel Case Concatenation rules (LC3)

The name of the concept must define the concept clearly with no chance of misunderstanding its content. The following rules will generally apply:

- The concept names MUST be based on appropriate presentation labels.
- The first character of the concept name MUST not be an underscore (_).

- The first character of the concept name **MUST** be capitalised.
- Words that do not add meaning **SHOULD** be left out. For example: a, an, in, on, at, where, that, which.
- Connective words in the label **MAY** be omitted from the concept name to make names shorter. Examples of English connective words include (but are not limited to) the following: the, and, for, which, of, a.
- Special characters **MUST** be omitted from the concept name, including the following:
() * + [] ? \ / ^ { } | @ # % = ~ ` ; : , < > & \$ € .
- Concept names **MUST** be limited to 256 characters.

c) Concepts **MUST be unique**

Two or more concepts **MUST NOT** share the same concept name. If they do, then uniqueness **SHOULD** be accomplished by either appending a distinguishing suffix or prefix.

d) Nature of concepts **SHOULD be defined as a suffix**

Standard suffixes that **SHOULD** be used for naming non-dimensional concepts:

- Only one suffix is to be used for this purpose.
- “Abstract” to be used for all abstract concepts.
- “TextBlock” to be used for all string concepts representing text blocks.
- “Percentage” suffix is added to concepts that use the Percentage data type.

- “Enumeration” suffix is added to concepts that use pre-defined strings.

Standard suffixes that SHOULD be used for naming dimensional concepts:

- “Table” to be added to concepts representing the hypercube level. The name of the Table is defined in plural, e.g. StandardsProtocolsOrMethodologiesTable. All “Table” elements should be in the hypercubeItemType substitution group and have their type set to string and their period type set to duration.
- “Axis” to be added to concepts representing the dimension level of an explicit or typed dimension, e.g. StandardsProtocolsOrMethodologiesAxis. All “Axis” elements should be in the dimensionItemType substitution group and have their type set to string and their period type set to duration.
- “Member” to be added to concepts representing the member level of a dimension, e.g. GreenhouseGasProtocolCorporateAccountingAndReportingStandardRevisedEditionMember. All “Member” elements should have their period attribute set to duration, abstract set to “true” and their type set to domainItemType.
- “LineItems” (plural) to be added to concepts that serve as placeholders which group concepts within a hypercube. The name of the Line Items is defined in plural, e.g. InformationAboutStandardsProtocolsOrMethodologiesUsedToCollectActivityDataAndCalculateScopeOneAndTwoEmissionsLineItems.

7.2.4.2 Label naming convention

Labels are provided in the taxonomy to minimise the need to use reference materials and to ensure that the user of the taxonomy is using the correct concept. A label should provide a concise but complete description of the concept.

The following goals SHOULD be achieved when adding labels to concepts:

- Each label describes the meaning of a concept;
- A label SHOULD facilitate the ability to locate a concept quickly;
- A label SHOULD be meaningful, recognizable, consistent, and easy to read;
- A label SHOULD be unique - users of the taxonomy do not need to refer to the concept name to be sure of its meaning.

e) Labels SHOULD NOT contain certain special characters.

The following characters should generally be avoided in creating concept labels:

? | > < : * " + ; = . & ! @ # { } \

e.g. do not use

“Scope 3: Purchased goods and services [member]”

Common exceptions that are allowed are “scope1+2” and “scope 1+2+3”.

Characters that are allowed are:

A-Z, a-z, 0-9, (,), comma, -, ‘, space, [], /

e.g. use

Scope 3, purchased goods and services [member]

The use of “/” should be avoided and substituted with the expression “or”.

Labels SHOULD be concise, follow established terminology and avoid being excessively descriptive. The following rules SHOULD be applied:

- All abstract elements grouped under an Extended Link Role (ELR) SHOULD start with “Disclosure of...”. The exception is “General information about the report”.
- All abstract elements used to group dimension information SHOULD start with “Disclosure of...”.
- “LinItems” elements SHOULD NOT start with “Disclosure of...”. “LinItems” that are rooted under an “Abstract” element SHOULD start with “Information about...”. “LinItems” that are rooted in other “LinItems” SHOULD NOT start with “Information about...” and SHOULD use alternative forms such as “Description of...” or “Reason for...”.
- The formulation “[label]”, comment” SHOULD be used instead of alternative formulations such as “Comment on” or “Comment about”.

There are various accepted ways to spell some terms, thus the following list of terms should be used in the Climate Change Reporting Taxonomy:

- “scope 1 + 2” SHALL be used to mean values that aggregate the emissions from the two scopes;
- “scope 1 and 2” SHALL be used for disclosures where values need to be presented separately for the two scopes;
- “CO2e” SHALL be used to mean carbon dioxide equivalent;
- “tonne” SHALL be used to mean a “metric tonne”.

Labels SHALL start with a capital letter and SHALL NOT use upper case, except for proper names and abbreviations

Example of proper use is

Scope 3, purchased goods and services [member]

IPIECA's Petroleum Industry Guidelines for reporting GHG emissions, 2003 [member]

Example of improper use is:

Scope 3, Purchased goods and services [member]

Kyoto Protocol Greenhouse Gases [member]

f) Dashes SHALL NOT be used in labels where commas can be used instead.

For example, DO NOT use 'Verification or assurance underway but not yet complete - the first year it has taken place [member]', but rather use 'Verification or assurance underway but not yet complete, the first year it has taken place [member]'.

An allowed example of a dash would be "Guinea-Bissau [member]".

Numbers SHOULD be expressed as text when less than 10, with the exception of reference to Scopes 1, 2 and 3

The expression of a number is a matter of judgement. The following rules for numbers SHOULD be considered when using numbers in labels:

- Exact numbers one through nine should be spelt out, except for percentages and numbers referring to parts of a book (for example, '5 per cent', 'page 2').
- Numbers of 10 or more should be expressed in figures.

- Exceptions are mentioned to Scopes 1, Scope 2 and Scope 3 concepts, as defined in the Greenhouse Gas Protocol.

g) Labels SHALL NOT have leading spaces, trailing spaces or double spaces.

h) Labels SHALL NOT have spaces also between square brackets, e.g.

[led disclosure][text block]

and not

[led disclosure] [text block]

Certain adjectives and prepositions used in labels SHOULD appear before or after the noun and be separated by a comma

For example, “scope1 and 2” or “scope 3,” or the mention to “gross” will be constructed for all non-abstract elements, in the following way:

{other} {noun}, {scope}, {gross}

An example of this would be:

“Disclosure of boundary used for greenhouse gas inventory, scope 1 and 2”

“Emissions target, comment”

“Emission value, gross”

i) Labels SHOULD avoid defining what they do or do not include.

For example, “General environmental regulations, including planning permissions [member]” SHOULD be avoided. What an item includes or excludes should be provided in the definition of the concept or the calculation linkbase.

The label component related to XBRL and not to Climate Change Reporting SHALL be placed between square brackets ‘[]’ at the end or beginning of the label

The component of labels placed in square brackets provides XBRL-related information that does not influence the information provided. For example:

- [08000000] Disclosure of emissions data;
- Boundary disclosure [abstract]

The following rules SHALL apply to standard labels and aligned with the naming convention

Abstract elements used to organize the taxonomy:

- SHALL append the word “[abstract]”;
- Abstract elements that are nodes of line items SHALL append the word “[line items]”;

Dimension elements:

- that are tables SHALL append the word “[table]”;
- that are axes SHALL append the word “[axis]”;

- that are domain members SHALL append the word “[member]”.

Non-abstract, reporting elements include:

- textBlockItemType concepts SHALL append the word “[text block]” to the label;
- percentItemType concepts SHALL append the word “[percentage]” to the label;
- booleanItemType concepts SHALL append the word [flag] to the label;
- StringItemType type concepts SHALL append in the label:
 - The word [status] is used when the reported fact resembles a status, e.g. “Absolute emissions trend comparatively to last reporting year [status][lead disclosure]”, which can have as status as “increase” and “decrease”; or “Provision of incentives for management of climate change issues and attainment of targets [status] [lead disclosure]”, which can have a status of “Incentives provided” and “Incentives not provided”;
 - The word [enumeration] when the reported fact makes use of a set of pre-defined string values.
- dateItemType concepts SHALL have in the label the word “date”;
- gYearItemType concepts SHALL have in the label the word “year”;
- integerItemType concepts SHALL have in the label the word “number”;
- massItemType concepts SHALL append the word [CO2] or [CO2e] to the end of the label.

7.2.4.3 Extended Link Roles (ELR)

j) Role definitions SHALL start with the ordering number.

For improved sorting of the extended link roles (ELR), the definitions of the ELRs SHALL start with an eight-digit number. The numbers allow sorting of the ELRs according to the structure of greenhouse gas reports, for example: “[010000000] Disclosure of governance”.

The numbering of presentation link ELRs closely follows the organisation of the information requested by the CDP Questionnaire, namely the numbering of pages in its online response system.

The numbering of definition link ELRs also relate to the organisation of the CDP information request and the numbering of the pages and questions in the online response system. Thus, the first two digits relate to the CDP page where that dimensional structure is first used and the third digit will relate to the order in which they are presented on the CDP page. In exceptional cases, some ELRs might use lettering such as “a” or “b”. In these situations, a previous ELR which has already modelled the concept can present an alternative (and simpler) view of the same structure.

7.2.5 Taxonomy validation

The following tests were conducted to make sure that the XBRL taxonomy complies with the XBRL specification and filing rules.

- **Tests of conformity with the standards of XML 1.0.**

Specific validation software has been employed, together with analysis and comparison with the standards defined by the W3C.

- **Tests of conformity with the XBRL specification.**

Various validators on the market have been employed, and a visual inspection has been made of the files in plain text (see "Software employed in the validations" below).

- **Tests of arithmetic restrictions.**

XBRL validation tests have been conducted against the arithmetic restrictions defined in the formula linkbase.

- **Tests of conformity with FRTA.**

Tests of adaptation to the FRTA 1.0. recommendation have been conducted. However, we accept failure of the following 'Must Errors' due to the avoidance of certain rules applicable to the use of Formulas 1.0, Extensible Enumerations 1.1, Dimensions 1.0 since they are not applicable to non-financial reporting frameworks:

- 3.1.1. A linkbase MUST NOT include any link elements (simple, resource, extended, or arc) not in an XBRL module or in the XBRL 2.1 Specification.
- 3.1.11. The role URI in a roleType element MUST be an LRR approved role or begin with the same scheme and authority parts as the target namespace of the taxonomy schema where it appears.
- 3.1.2. An arc MUST have only its standard or LRR approved arc role.
- 3.1.7. All arcs within an extended-type link MUST have the same arc role.
- 4.2.1. A schema document MUST contain only declarations of reference parts OR declarations of concepts, roles and arc roles OR declarations that are not concepts and not reference parts.
- 4.3.2. Each unique taxonomy schema target namespace must have one (and only one) namespace prefix of one to twelve characters, which will be its recommended namespace prefix.

7.3 Conclusions

This chapter presents the key contributions I made in the area of technology standards to accelerate the adoption of environmental data into valuable information for decision-making. As outcomes, an Horizon H2020 report was delivered and an XBRL Taxonomy for environmental information was published at the CDP website. Both works have generated an impact on the Spanish companies with more than 500 employees to disclose the environmental data required by the European Non-financial reporting directive. This taxonomy has represented an evolution

of the XBRL specification to be capable to represent environmental data. In this respect, the Enumeration Specification³¹ was evolved to facilitate a better analysis of qualitative information based on the feedback I provided to XBRL International using this specification for the CDP taxonomy.

³¹ XBRL Enumeration specification: <https://specifications.xbrl.org/work-product-index-extensible-enumerations-extensible-enumerations-2.0.html> [Accessed August 2019]

Chapter 8. Adopting Semantic Technologies for Effective Corporate Transparency

In Chapters 5 and 6 we identified the need to link factors regarding non-financial performance of corporations into the decision-making processes of investors and other stakeholders. To do this, we need to develop better ways to access and analyse corporate social, environmental and financial performance information, and to link together insights from these different sources. Such sources are already on the Internet in non-structured and structured data formats, many of them in XBRL (Extensible Business Reporting Language). This chapter is trying to answer our third objective “How to enable the use of environmental data in combination to financial information from companies ?”.

This chapter proposes a solution to enhance corporate transparency to increase the adoption of financial and non-financial data given the current adoption of XBRL and the new opportunities offered by Linked Data. We present (1) a methodology to formalise XBRL as RDF using Linked Data principles, and (2) demonstrate its usefulness through a use case connecting and making the data accessible. At the end, a new transparency model is envisioned, with improved corporate data to promote sustainable economic growth.

This work was presented and published at 14th European Semantic Web Conference. (pp 655-670). Published by Springer International Publishing. DOI: 10.1007/978-3-319-58068-5_40 . To clarify, that this work and examples are targeted to Semantic Web experts given the audience of this conference.

Mora-Rodriguez, M., Auguste-Atemezing, G. and Preist, C. (2017). Adopting Semantic Technologies for Effective Corporate Transparency. 14th European Semantic Web Conference. (pp 655-670). Published by Springer International Publishing. DOI: 10.1007/978-3-319-58068-5_40

Paper written by Maria Mora, with revisions from Chris Preist and Ghislain Auguste (Modeca Lab Research). Presented by Maria Mora at the 14th European Semantic Web Conference (Portoroz, Slovenia). June 2017. Presentation available here: <https://www.slideshare.net/MariaMora69/adopting-semantic-technology-for-effective-corporate-transparency>

8.1 Introduction

Transparency is increasingly used as a means for holding organisations to account, both in the public and private sector. In this study, we focus primarily on the latter: the role of transparency and open data to promote good governance and trust between the private sector and its diverse stakeholders. The main tool for this is corporate reporting—the self-disclosure of information by a company to a set of stakeholders in a well-defined and routine way. Such stakeholders are primarily, although not exclusively, existing and potential investors. These stakeholders need to satisfy themselves as to the financial performance and good governance of the company they invest in. Many governments require such reports to be in specific formats, and are increasingly making them publically available through open data initiatives, such as the EDGAR program of the U.S. Security and Exchange Commission and the data repository of the Spanish Security Exchange Commission (CNMV).³²

Such data is often submitted and made available in XBRL (Extensible Business Reporting Language), an XML format adopted to make corporate data more standardised and exchangeable. XBRL is currently in use in more than 60 countries, implemented by over 100 regulators that cover 10 million companies worldwide.

In addition to mandatory reporting, there are a number of voluntary initiatives encouraging corporations to disclose information regarding their performance in areas of economic, social and environmental impact. These include the CDP (Carbon Disclosure Project), GRI (Global Reporting Initiative), SASB (Sustainability Accounting Standards Board) and the IIRC (International Integrated Reporting Council). These are often international non-governmental

³² CNMV portal: <http://www.cnmv.es/ipps/Default.aspx> [Accessed August 2019]

organisations, with representatives of both corporations and other stakeholders including investors, academics, environmental NGOs and policymakers in their governance structures. Some of these promote the disclosure and use of their data through XBRL, in a similar way to governmental open data initiatives.

Such initiatives encourage engagement with this data: data journalists can investigate corporate behaviour; investors can integrate future risks associated with factors such as climate change in their assessment of companies; companies can benchmark themselves against others; academics can explore wider trends and correlations in financial performance and non-financial behaviours.

However, such transparency alone is not enough to support companies and their stakeholders' decisions and actions. There are several barriers such as:

- The lack of easily accessible corporate data to quickly and accurately inform the management about pertinent issues during decision-making-processes.
- Unfamiliarity in how sustainable aspects have an impact on financial outcomes and vice versa.
- Inadequate levels of integration of financial and non-financial information within the internal performance, strategy and operational frameworks of an organisation.
- A dearth of consistency, comparability, reliability and clarity of climate change information emerging from organisations globally. Standardisation and mainstreaming of disclosures needs to be facilitated.

The exposure of XBRL reports as open data is a first step to overcome some of these barriers. XBRL allows access to standardised data in an open format about corporate financial environmental sustainability. However, XBRL offers limited interconnection between them. In particular, XBRL exhibits the following weaknesses:

- It is primarily structured around documents and entities rather than data, making links between data elements difficult.
- The same data structures can be modelled in different ways, generating different technical implementations.

The Open Information Model (OIM) is being developed by the XBRL community in an effort to increase the adoption of existing XBRL data by reducing the heterogeneity of the XBRL reports generated by different modelling practices and enabling a better integration with Information Systems. OIM aims to facilitate the serialisation of XBRL reports in CSV, JSON and XML formats. In this study, we use the Open Information Model as a template for extracting linked open data from XBRL documents, thus allowing them to be combined more easily and to be used alongside other open data sources to enrich analysis by stakeholders. We demonstrate this by working with data from the Spanish Security and Exchange Commission (CNMV) alongside data from CDP.

Our study makes a contribution in the field of Semantics and Corporate Transparency by more effectively integrating data sources using XBRL, the most common corporate standard, into the Linked Data environment. We show that this in turn makes contributions in the following areas:

- The generation and integration of Linked Open Government Data made available from government sources (CNMV) in a different format (XBRL).
- Provenance and accountability of companies, showing how financial and non-financial data from different sources can be linked using our approach to hold companies more environmentally accountable.
- Trust, data traceability and fact checking of corporate data sources, enabled by using the proposed ontology together with SPARQL to facilitate cross-checking of corporate data with alternate sources.

This study is organised as follows: in section two and three, we provide background on XBRL and prior work connecting XBRL with Linked Data formats. Section four presents a methodology to transform XBRL to RDF following Linked Data principles and the Open Information Model.³³ We describe a case study implementing our proposal, using financial

³³ <http://www.xbrl.org/Specification/oim/PWD-2016-01-13/oim-PWD-2016-01-13.html>

and environmental data in XBRL format from the Spanish companies Repsol S.A. and Amadeus IT Group, published by CNMV and CDP respectively. Having XBRL data in RDF format, we proceed to link with other data sources from LOD using LIMES³⁴ as a Link Discovery Framework. We then make the RDF data accessible and queryable via SPARQL endpoint, and we proceed to evaluate some queries to show the potential benefits for data users. Section six presents and discusses the results. The study closes with conclusions and lessons learned that, if addressed by the XBRL and Linked Data community, would further promote transparency.

8.2 XBRL fundamentals relevant for this chapter

XBRL aims to overcome the limitations of traditional and paper-based disclosures [104]. Through data standardisation in an open digital format, XBRL can help enhance data quality and data analysis through:

- An open mechanism to represent contextualised business facts under defined business requirements (presentation, period, legal references, calculation) and data quality;
- Enabling data-driven decision management;
- Improved accessibility and integration of the information to any application or management process;
- Standardised validation and comparability of information.

For financial disclosure regimes like the International Financial Reporting Standards (IFRS) and US Generally Accepted Accounting Principles (US GAAP), or non-financial disclosure regimes like the CDP and GRI, a single XBRL taxonomy is created. The taxonomy is where the rules and data definitions are organised. It is comprised of a set of elements (i.e., Key Performance Indicators and narratives) and all the presentation, calculation, labels in different languages, and standard logic rules (linkbases) which provide semantic meaning. The taxonomy also includes mechanisms for defining reporting requirements in a multi-lingual context; this permits filers to use their language of choice while also allowing consumers to review it in their own.

³⁴ <http://aksw.org/Projects/LIMES.html>

Once created, the XBRL taxonomy is published online. Then, for a given firm, software can be used to create an XBRL instance (the report itself), containing facts and figures for a certain period (Figure 28). The XBRL instance can be checked against the taxonomy by all parties (reporting entity, a regulator, or even the public) in order to guarantee its data quality and reliability as the taxonomy contains data quality checks that any XBRL engine can validate. The validation rules supported in XBRL allow a good level of data quality, from basic rules to validate data types (number, text, precision), to more complex rules relating to elements that have been disclosed. For example, rules can be implemented to check if a breakdown of carbon emissions is equal or not to the total emissions reported, or a CO2 intensity figure (tCO2/revenue) is actually in line with revenue and emissions figures reported.

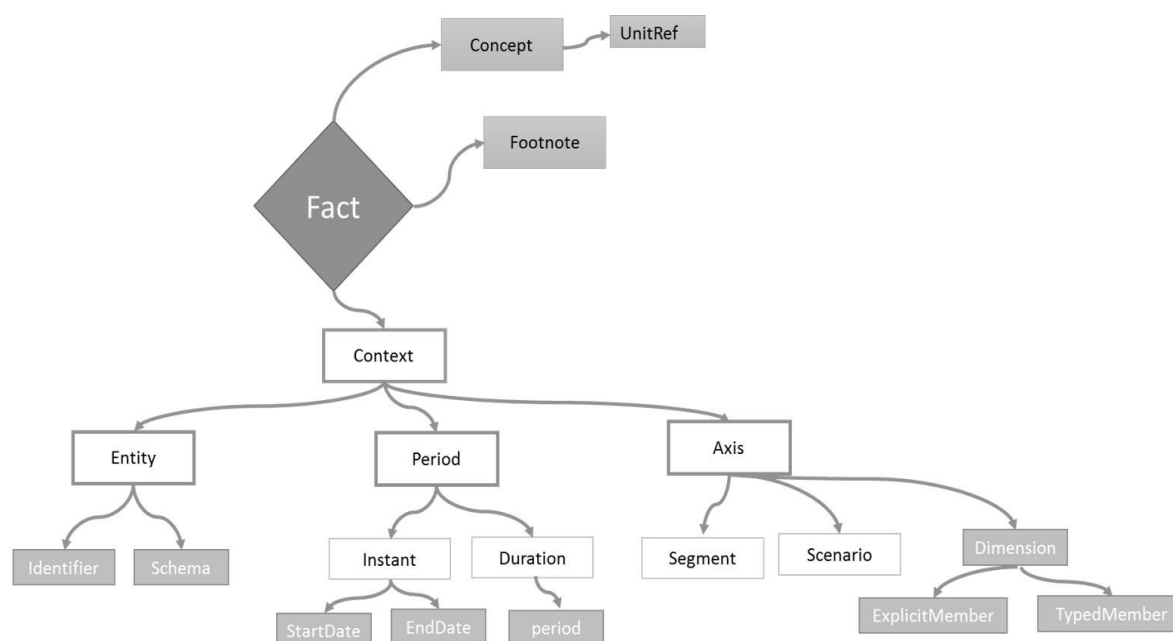


Figure 28. Fact representation in XBRL

The XBRL core specification (XBRL 2.1) has evolved since its creation in 2001 to enrich its dimensional data representation and validation rules. The XBRL Dimensions 1.0³⁵ in 2006 and Formulas 1.0³⁶ in 2009 provide optional incremental syntax to the core specification.

³⁵ <http://www.xbrl.org/specification/dimensions/rec-2012-01-25/dimensions-rec-2006-09-18+corrected-errata-2012-01-25-clean.html> [Accessed August 2019]

³⁶ <http://www.xbrl.org/wgn/xbrl-formula-overview/pwd-2011-12-21/%20xbrl-formula-overview-wgn-pwd-2011-12-21.html> [Accessed August 2019]

Dimensions 1.0 enriches rules and procedures for constructing dimensional taxonomies and therefore instance documents. Taxonomies using Dimensions specification can define new dimensional contexts, specifying valid values ("domains") for dimensions, using a mechanism called hypercube to define which dimensions apply to which business concept. There are two types of dimensions:

- Explicit dimensions: These have a fixed number of dimension members. For example, in a two-dimensional table, the number of row and columns are known.
- Typed dimensions: the number of dimension members is unknown. For example, in a two-dimensional table, it means that the number of columns is known, but the number of rows depends on user reporting needs.

However, the Dimensions specification in XBRL 2.1 has certain limitations. It does not fully support calculation rules (defined in calculation linkbase); calculations cannot be executed across different contexts. In other words, in a simple two-dimensional table, calculations can be executed by columns and not by rows. In part, these limitations, and the need to have strong validation capabilities in XBRL, resulted in the Formulas 1.0 specification in 2009. This module enhances the XBRL validation capabilities, using XPath to validate instances and to calculate new XBRL facts.

XBRL has evolved to support both global regulatory environments and emerging domains such as sustainability reporting. The flexibility needed to do this has resulted in difficulties with regard to standardisation. Primarily, this is because of the diversity of technical implementations produced by different modelling practices during the taxonomy development phase. Though XBRL is a standard, it offers different ways to model data structures. For example, tables can be represented in XBRL using tuples or dimensions. This is true of the two taxonomies we use in our work later in the study: the CNMV taxonomy represents financial facts as items and tuples, and only makes use of the XBRL 2.1 core specification. Items are facts holding a simple value represented by a single XML element with the value as its content and period, and information about its reporting entity as a context attribute. An example item could be equity in the last quarter. Tuples are facts holding multiple values, and they are represented by a single XML element containing items or other

tuples. For example, preferred stock is always defined by the combination of different stocks. Thus, preferred stock is a tuple defined by two items: the preferred stock-nominal value and the preferred stock-shares authorised. The CDP taxonomy, on the other hand, represents (environmental) facts as simple items and dimensional structures and makes use of XBRL 2.1 core and Dimensions 1.0 specifications. Dimensional structures are represented as items where context attributes also include dimensional XML elements.

The new Open Information Model (OIM) specification proposes an independent XBRL model to represent XBRL business facts; it focuses on XBRL instance documents instead of taxonomy definitions. This overcomes the taxonomy modelling difficulties and keeps the value and context of the data. However, the semantic richness of the taxonomy disappears, such as advanced validation rules, human labels in different languages and how the data should be presented.

In our study, we propose an ontology based on OIM specification to transform XBRL data to RDF, using as a case of study financial and environmental company data from CNMV and CDP.

8.3 Transforming XBRL into linked data

8.3.1 Lightweight vocabulary for XBRL (XBRLV)

We developed a lightweight ontology using the Web Ontology Language (OWL) based on XBRL standard (XBRL 2.1 and Dimensions 1.0). The goal of implementing a lightweight vocabulary for XBRL is threefold: (1) Easy identification of the key concepts of the XBRL standard; (2) Reuse of existing vocabularies to describe XBRL datasets; (3) Enrichment and linking of data with relevant datasets in the Linked Open Data cloud.

Unlike previous efforts, we based our ontology proposal on OIM, which is a syntax-independent model of the content of an XBRL report instead of taxonomy definition. OIM defines four components:

Namespace: Representation of XML namespace prefixes.

DTS Reference: Reference to XML documents and schema linked to an XBRL report.

Report: Top-level component that encapsulates the data of an XBRL report.

Fact: Representation of a business fact in an XBRL report. As explained in section one, a fact can be a simple item, a tuple fact or a dimension. All facts have the following common properties (ID aspects and footnotes). ID is a unique identifier; aspects are properties which represent the entity and period that a business fact refers to (oim:entity and oim:period). Examples of facts include:

- the unit of measure, such as "USD" and "MWh" (oim:unit);
- the reporting item (oim:concept);
- tuples definition, which represents a grouping container for other facts, dimensional structure, axis and members;
- the footnotes of the fact (oim:footnotes).

Our XBRL-Lightweight (XBRL-L) ontology is composed of 15 classes, 12 object properties and 12 data properties, with DL expressivity: ALC(D). The ontology follows best practice in the semantic web by reusing existing ontologies to improve data interoperability (Breslin *et al.*, 2009), through the Linked Open Vocabularies initiative (LOV³⁷).

We provide a hierarchical structure following OIM, mapping XBRL components to Semantic web vocabularies. The class Report is a subclass of schema:Report³⁸. The class Fact is used to represent XBRL business facts (items, tuples and dimensions) that in turn refer to entity (hasEntity), concept, period, scenario, value and footnote modelled as object properties. In the case of numeric and currency values, the number of decimals and unit type are also represented. These properties are enough to represent a simple XBRL item.

³⁷ <https://lov.okfn.org/> [Accessed August 2019]

³⁸ <http://schema.org> [Accessed August 2019]

As required from XBRL, a fact can hold multiple values in the form of tuples and dimensional structures. For that purpose, we defined the `hasTuple` and `hasDimension` properties as part of the `Fact` class. They point out to `Tuple` and `Dimension` classes respectively. `Tuple` class is composed by `concept` and `hasTuple` properties. The latter means a tuple can be embedded as part of another tuple, defining the context of the main item. `Dimension` class is composed of `Axis` and `Member` properties, representing the axis and member per axis which define the context of the main item. We decided not to differentiate whether the axis is explicit or typed, as our model is focused on instance documents instead of taxonomy definition. It means that as we are working with XBRL reports, the dimensional members are defined.

Unlike the OIM model, we decided to define `Entity` as a class instead of an object property of the class `Fact` that links to the schema identifier of the entity that is part of the XBRL report. In that way, we extend the class `rov:RegisteredOrganization`³⁹, (1) keeping the correspondence with well-known vocabularies and (2) allowing the full information of the reporting firm facilitating the discovering link process. Figure 29 shows classes and relationships defined in the XBRL ontology, which is available at <https://w3id.org/vocab/xbrl#>.

³⁹ <http://www.w3.org/ns/regorg#> [Accessed August 2019]

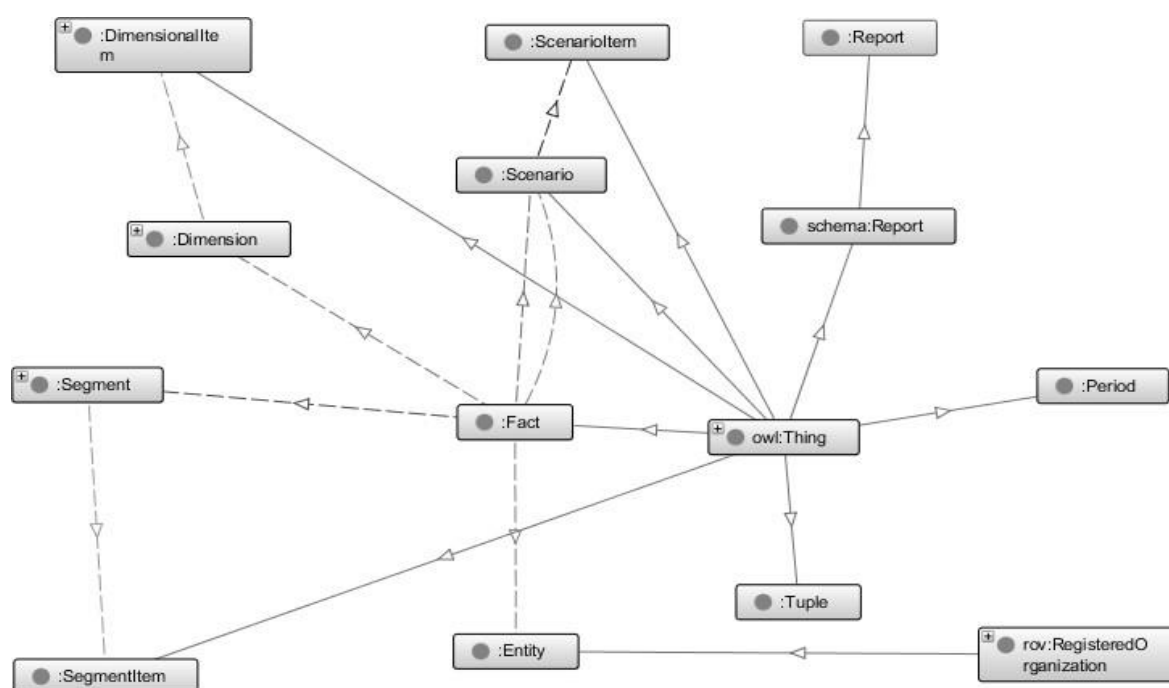


Figure 29. Classes and relationships XBRL Ontology

8.3.2 From XBRL data to Linked data

In the next step we demonstrate how XBRL data can be mapped to the ontology and how it can be published using Linked Data principles. We demonstrate our method using financial and environmental XBRL data from the Spanish companies Repsol and Amadeus IT Group, chosen as those that are published in both CNMV and CDP⁴⁰. For the transformation process, we developed a script in Python (<https://goo.gl/VggJQZ>) to transform the XBRL reports using JSON files generated by the Arelle⁴¹ open source platform. Note that the JSON data used as our input is generated according to OIM.

A simple fact in XBRL (Figure 30), is the representation of a concept (cdp:IntroductionCompany) and its value, where the context consists of the period, unit and information about the reporting company (entity).

⁴⁰ The CDP data used is publicly available, but not yet in XBRL format. This is currently only available internally to CDP, and made available to this project.

⁴¹ <http://arelle.org/> [Accessed August 2019]

A tuple fact in XBRL (Figure 31) represents facts with multiple values. In this case, the concepts ifrs-gp:IntangibleAssetsNet and ifrs-gp:GoodwillNet are the elements that compose the tuple ipp-gen:BalanceIndividual. The CNMV allows reporting companies to use xbrli:scenario element to determine if the value is part of an individual or consolidated financial statement. Hence in CNMV reports we find the same tuple hierarchy and concepts linked to two different contexts: consolidated and individual. Currently, the scenario and segment elements, in a non-dimensional domain, such as the CNMV reports, are not considered either by OIM and Arelle when transforming XBRL into JSON⁴². Our ontology considers both.

<pre> <http://data.mondeca.com/id/fact/f001> a ns1:Fact ; ns1:hasEntity <http://data.mondeca.com/id/entity/000084 7838> ; ns1:concept cdp:IntroductionCompany; ns1:period [ns1:startPeriod "2015- 01-01"^^xsd:date ; ns1:endPeriod "2015-12- 31"^^xsd:date] ; ns1:value "Repsol is international integrated Oil and Gas Company whose main activity consists of the upstream and downstream business." ; </pre>	<pre> <xbrli:context id="ctx_00"> <xbrli:entity> <xbrli:identifier scheme="http://www.cdp.net/CIK">0000847838</xbrli:identifie r> </xbrli:entity> <xbrli:period> <xbrli:startDate>2015-01-01</xbrli:startDate> <xbrli:endDate>2015-12-31</xbrli:endDate> </xbrli:period> </xbrli:context> <cdp:IntroductionCompany contextRef="ctx_00">Repsol is international integrated Oil and Gas Company whose main activity consists of the upstream and downstream business.</cdp:IntroductionCompany> </pre>
--	--

Figure 30. RDF representation of a simple XBRL fact

⁴² The OIM working group and Arelle's authors were informed about the lack of segment and scenario representation in a non-dimensional domain.

<pre> <http://data.mondeca.com/id/fact/f0004> a ns1:Fact ; ns1:concept ifrs-gp:IntangibleAssetsNet; ns1:entity <http://data.mondeca.com/id/entity/A7837472>; ns1:period [ns1:instant "2015-12-31"^^xsd:date] ; ns1:hasTuple <http://data.mondeca.com/id/tuple/t89493> ; ns1:unitRef < http://dbpedia.org/resource/EUR> ; ns1:value 94588000.0 ; ns1:decimals 0 . ns1:hasScenario <http://data.mondeca.com/id/scenario/s0001> . <http://data.mondeca.com/id/fact/f0004> a ns1:Fact ; ns1:concept ifrs-gp:GoodwillNet ; ns1:entity <http://data.mondeca.com/id/entity/A7837472>; ns1:period [ns1:instant "2015-12-31"^^xsd:date] ; ns1:hasTuple <http://data.mondeca.com/id/tuple/t89493> ; ns1:unitRef < http://dbpedia.org/resource/EUR>; ns1:value 350.0 ; ns1:decimals 0 . ns1:hasScenario <http://data.mondeca.com/id/scenario/s0001> . <http://data.mondeca.com/id/scenario/s0001> a ns1:Scenario ns1:scenarioItem [ns1:concept ipp:Modelo; ns1:value "GE"] ns1:scenarioItem [ns1:concept ipp:Apartado; ns1:value "Individual"] <http://data.mondeca.com/id/tuple/t89493> a ns1:Tuple ; ns1:concept ipp-gen:BalanceIndividual; ns1:hasTuple <http://data.mondeca.com/id/tuple/t89494> </pre>	<pre> <xbrli:context id="S22015_A-78374725_ici"> <xbrli:entity> <xbrli:identifier scheme="http://www.cnmv.es/xbrl/ipp/A- 78374725">REPSOL, S.A</xbrli:identifier> </xbrli:entity> <xbrli:period> <xbrli:instant>2015-12-31</xbrli:instant> </xbrli:period> <xbrli:scenario> <ipp:Modelo>GE</ipp:Modelo> <ipp:Apartado>Individual</ipp:Apartado> </xbrli:scenario> </xbrli:context> <xbrli:unit id="euro"> <xbrli:measure>iso4217:EUR</xbrli:measure> </xbrli:unit> <ipp-gen:InformacionFinancieraSeleccionada> <ipp-gen:BalanceIndividual> <ifrs-gp:IntangibleAssetsNet decimals="0" contextRef="S22015_A-78374725_ici" unitRef="euro">94588000</ifrs- gp:IntangibleAssetsNet> <ifrs-gp:GoodwillNet decimals="0" contextRef="S22015_A-78374725_ici" unitRef="euro">350</ifrs-gp:GoodwillNet> </ipp-gen:BalanceIndividual> </ipp-gen:InformacionFinancieraSeleccionada> </pre>
---	---

Figure 31. RDF representation of a tuple fact

Dimensional facts in XBRL (Figure 32) can also represent multiple values. For example, the concepts `cdp:EmissionValueGross` and `cdp:Scope` are linked to the same Axis `cdp:TotalEmissionDataAxis` and related member(`cdp:GreenhouseInventoryBoundariesID`). This structure allows disclosing the total Emissions gross values per type of scope (Scope 1, Scope 2 locationbased, Scope 2 market-based and Scope 3). Here the value of 21068516 CO2e corresponds to the Scope 1 emissions.

<pre> <http://data.mondeca.com/id/fact/f0005> a ns1:Fact ; ns1:concept cdp:EmissionValueGrossCO2e ns1:entity <http://data.mondeca.com/id/entity/00008478 38> ; ns1:period [ns1:startPeriod "2015-01- 01"^^xsd:date; ns1:endPeriod "2015-12- 31"^^xsd:date] ; ns1:hasDimension <http://data.mondeca.com/id/dimension/d0001 > ; ns1:unitRef <http://dbpedia.org/page/Carbon_dioxide_equ ivalent>; ns1:value 21068516; ns1:decimals 0 . . <http://data.mondeca.com/id/dimension/d0001 > a ns1:Dimension ; ns1:dimensionItem [ns1:axis TotalEmissionDataAxis; ns1:value "id01"] </pre>	<pre> <xbrli:context id="ctx_8_2_2015_id01"> <xbrli:entity> <xbrli:identifier scheme="http://www.cdp.net/CIK">0000847838</xbrli:identifier> <xbrli:segment> <xbrldi:typedMember dimension="cdp:TotalEmissionDataAxis"> <cdp:GreenhouseInventoryBoundariesID>id01</cdp:GreenhouseInvento ryBoundariesID> </xbrldi:typedMember> </xbrli:segment> </xbrli:entity> <xbrli:period> <xbrli:startDate>2015-01-01</xbrli:startDate> <xbrli:endDate>2015-12-31</xbrli:endDate> </xbrli:period> </xbrli:context> <xbrli:unit id="CO2e"> <xbrli:measure>cdp:CO2e</xbrli:measure> </xbrli:unit> <cdp:EmissionValueGrossCO2e decimals="0" contextRef="ctx_8_2_2015_id01" unitRef="CO2e">21068516</cdp:EmissionValueGrossCO2e> <cdp:Scope contextRef="ctx_8_2_2015_id01">cdp- enum:Scope1</cdp:Scope> </pre>
---	--

Figure 32. RDF representation of a dimensional fact

During the transformation of the CNMV report, we found certain XBRL elements with content about persons and activities that belong to the imported XBRL taxonomy called Data of General Identification (DGI)⁴³. We map these using Friend Of a Friend (FOAF) vocabulary ⁴⁴ (Figure 33), in line with best practice of reusing existing vocabularies in specific contexts to increase the level of interoperability.

<pre> <http://data.mondeca.com/id/xbrl/person/016 06701J> a ns5:Person ; ns7:identifier "01606701J" ; ns6:role "Consejero" ; ns5:name "D. Javier Echenique Landiribar" ns2:codeID "CIF" . </pre>	<pre> <ipp-com:DetallePersonasResponsables> <dgi-est-gen:PersonName> <dgi-est-gen:FormattedName contextRef="S12015_A- 78374725_ici">D. Ángel Durández Adeva</dgi-est- gen:FormattedName> </dgi-est-gen:PersonName> <dgi-est-gen:Identifier> <dgi-est-gen:IdentifierCode> <dgi-lc-es:Xcode_IDC.CIF contextRef="S12015_A- 78374725_ici">CIF</dgi-lc-es:Xcode_IDC.CIF> <dgi-est-gen:IdentifierValue contextRef="S12015_A- 78374725_ici">01606701J</dgi-est-gen:IdentifierValue> </dgi-est-gen:IdentifierCode> </dgi-est-gen:Identifier> <dgi-est-gen:PositionType contextRef="S12015_A- 78374725_ici">Consejero</dgi-est-gen:PositionType> </ipp-com:DetallePersonasResponsables> </pre>
--	---

Figure 33. DGI representation in RDF using FOAF vocabulary

⁴³ https://joinup.ec.europa.eu/asset/data_of_general_identification/home [Accessed August 2019]

⁴⁴ <http://xmlns.com/foaf/spec/> [Accessed August 2019]

As in XML, XBRL Namespaces specifications do not need to reference a real location — they just need to be unique. However, in RDF the namespace URI must identify the location of the schemas. As certain XBRL namespaces from the CNMV reports were not valid, we had to store the schemas in our server and point out the namespaces to real locations. In many cases, we decided to map the XBRL units and currencies to well-known DBpedia links, connecting related data that were not previously linked. For example:

`<xbrli:measure>iso4217:EUR</xbrli:measure>` to <http://dbpedia.org/resource/EUR>

`<xbrli:measure>cdp:CO2e</xbrli:measure>` to
[http://dbpedia.org/page/Carbon dioxide equivalent](http://dbpedia.org/page/Carbon_dioxide_equivalent)

As XML, an XBRL document forms a tree structure ready to be consumed as a full report [111]. The move to data consumption requires the use of dereferenceable URIs to denote facts in a unique way, keeping its context. We use the following URI conventions to denote related facts and classes:

Fact: `http://data.mondeca.com/id/fact/f[0-9]*` → `http://data.mondeca.com/id/fact/f88557`

Entity: `http://data.mondeca.com/id/entity/(company identifier)` → `http://data.mondeca.com/id/entity/A-78374725`

Through the dereferenceable URIs, facts can be visualised using open source tools like LodLive⁴⁵. We provide an example here: <https://goo.gl/iFVE0B>.

8.3.3 Linking XBRL data to other data

If transparency is to be enabled, it is very important to convert the independent XBRL silos of information into pieces connected with existing Linked Data sources available on the web.

⁴⁵ <http://en.lodlive.it/> [Accessed August 2019]

For that purpose, we use LIMES, which is a tool that allows detecting similar Linked datasets. LIMES works by specifying the search criteria and the target endpoint to search in. Our search criteria is the company name contained in the Entity fact from the generated RDF. The DBPedia endpoint is the target source that we choose to gather the links, restricting the search by `sch:Organization`. LIMES requires a metric and acceptance condition setting a threshold value. We use the trigrams metric offered by LIMES to mapping correspondences between the `ns6:legalName` of our local RDFs and the `sch:Organization` from DBpedia. For the purpose of this study, we only accept results with a minimum of 0.90 level acceptance. The final results were the following URLs <http://dbpedia.org/resource/Repsol> and [http://dbpedia.org/page/Amadeus IT Group](http://dbpedia.org/page/Amadeus_IT_Group), included as a SameAs relationship in our local RDF files.

8.4 Validation

For validation purposes, we ran queries against the final ontology generated, evaluating its quality and accuracy by checking whether they contain enough information to cover three goals to promote effective transparency:

Data coverage: through better data contextualization.

Better data analysis: enabling cross-data-source analyses.

Data accuracy: facilitating data cross-checking contained in different sources.

For that, we first implemented an endpoint⁴⁶ using Apache Jena Fuseki⁴⁷, available here: <http://data.mondeca.com/dataset.html?tab=query&ds=/xbrl-data>. We used SPARQL (Simple Protocol and RDF Query language) because it allows us to express queries across diverse data. We conducted three queries with each of the two companies (Repsol and Amadeus IT Group) data. We illustrated each query below with one of the companies. Full results are available at the SPARQL queries provided. Figure 34 presents the final architecture.

⁴⁶ <http://data.mondeca.com/xbrl-data/sparql> [Accessed August 2019]

⁴⁷ <https://jena.apache.org/> [Accessed August 2019]

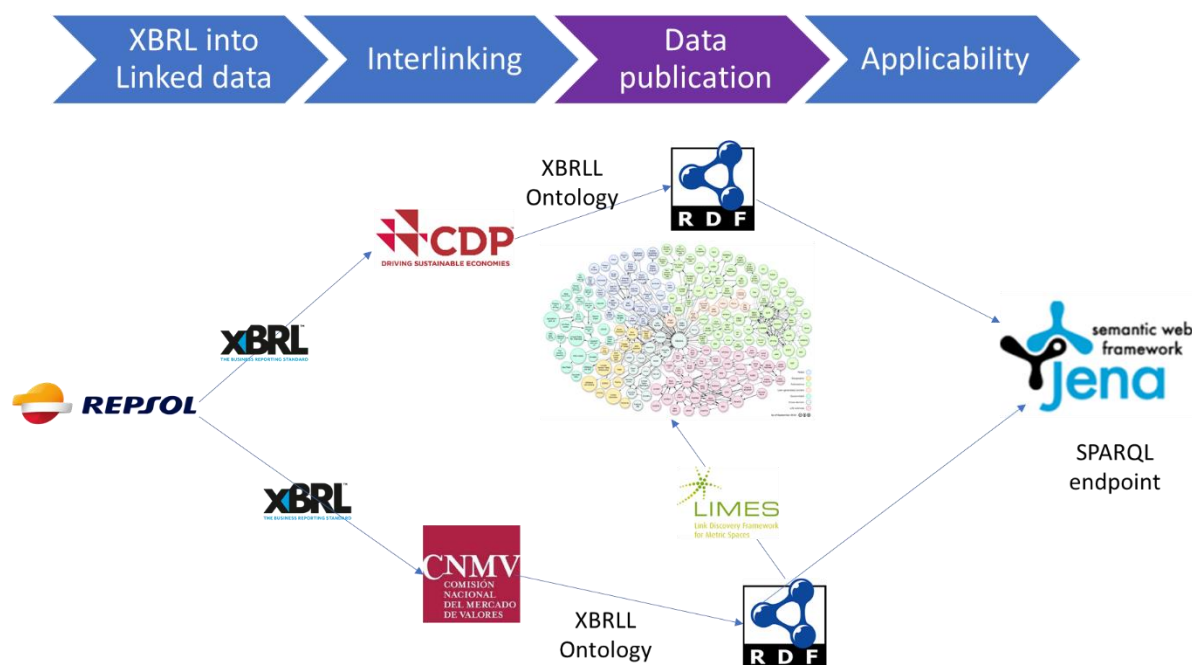


Figure 34. From XBRL to Linked data architecture. Inputs, outputs and technologies

Outputs of this work are presented below:

Goal 1. Data coverage using DBpedia

Question: What is the context of the company Repsol?

Data: Abstract, subsidiary and industry.

SPARQL query: <https://goo.gl/if8ydG>

Output: presented in Table 15.

URL	Abstract	Subsidiary	Industry
http://dbpedia.org/resource/Repsol	Repsol S.A. is an integrated global energy company based in Madrid, Spain. It carries out upstream and downstream activities	http://dbpedia.org/resource/Petronor	http://dbpedia.org/resource/Petroleum_industry

	across the globe. It has more than 24,000 employees worldwide.		
--	--	--	--

Table 15. Data coverage: Information about the context of Repsol S.A.

Goal 2. Cross data source analysis using CNMV and CDP data

Question: What was the emission intensity of Repsol in 2015?

Data: Scope 1 emissions (CDP) divided by Consolidated sales (CNMV).

SPARQL query: <https://goo.gl/7bIE9m>

Output: presented in Table 16.

CO2 Emissions	Consolidated sales	Emission intensity
21068516	39737000000	0.00053

Table 16. Data analysis: Emission intensity of Repsol S.A. in 2015

Goal 3. Data accuracy using DBpedia and CNMV data

Question: How reliable is the equity figure presented in DBpedia?

Data: Equity (DBpedia) and equity (CNMV) in the year 2013.

SPARQL query: <https://goo.gl/LGb53s>

Output: presented in Table 17.

Entity name	Equity(DBpedia) dbo:equity	Equity(CNMV)ipp- gen:PatrimonioNetoNiif	Difference
Repsol S.A	2.792E10	27920000000	0.0001
Amadeus IT Holding	€1,840.1 million@en	1840066000.0	-

Table 17. Data consistency: Reliability of Equity figure presented in DBpedia

8.5 Discussion

This study demonstrates that Linked Data can be used to integrate financial and non-financial data and can facilitate transparency among diverse stakeholders. Our work does this by converting corporate XBRL reports into RDF and linking them to other relevant financial and

non-financial data (e.g., environmental, DBpedia). A generic ontology to transform any XBRL report into Linked Data has been proposed, along with ways to resolve the lack of formal correspondences with well-known vocabularies. This solution overcomes the XBRL challenges related to the diversity of technical implementations produced by different modelling practices and so goes beyond prior related works.

We demonstrate that using Linked Data with well-adopted standards, such as XBRL, improves the interoperability and access to existing corporate datasets, as well as straightforward integration with related data in other formats. The validation demonstrates that the solution proposed offers three benefits for data users: increased data coverage, better data analysis and data consistency. These benefits are real because we have built a solution capable to cover the following three goals :Data coverage, Cross data source analysis and Data accuracy. Besides, the solution presented can be replicable and through the links provided and explained in section 8.5 can be run. Our intention with this work is to demonstrate that technically it is possible to interconnect and publish pieces of information from different domains using existing technologies and publication principles like Linked data and it is not just a theory. It is out of scope of this work to tackle how the combination of this information can drive specific business decisions.

The results in Table 17 present an interesting point for discussion. It shows that the DBPedia data (dbo: equity) lacks context and numeric precision. For example, there is no year associated with the equity figure nor consistent use of datatype. Amadeus equity is a string €1,840.1 million while Repsol equity, which has the same tag (dbo:equity), is a number. The data from CNMV in XBRL format does not have any of those problems.

Given these results, we believe that XBRL is a better format than RDF to standardise corporate information. However, it is less able to connect different data silos in various formats. For that, the publication of data using Linked Data principles is the most appropriate solution.

This study proposes the combination of both solutions, XBRL and Linked Data, to improve corporate transparency. Below, we enumerate a set of technical requirements to apply on

XBRL schemas, definitions and reports to converge towards a Linked Data approach. Adoption of these best practices in XBRL modelling would enhance interoperability and transparency of corporate data.

- Use of common data structures must be encouraged in XBRL taxonomies. For example, taxonomies such as DGI to represent common corporate information such as company name, unique identification number, activities and sectors. This would not only enhance the interoperability between XBRL data from different taxonomies but also ease the mapping process with well-known vocabularies in the Linked Data world.
- The use of namespaces notation that point out to real locations should be promoted in XBRL. This would ease the transformation of XBRL data into Linked Data, facilitating better inference mechanisms.
- Using dereferenceable URIs to denote and identify XBRL facts provides a way to access and link relevant information to those objects across the web. It enables better interoperability between data in XBRL format and other data sources.
- Reusing existing RDF data on units and currencies already published in LOD brings more contextual information than the current ISO and XBRL units reference.
- Using tools like LINES can help to increase the coverage of information by continuously integrating data sources.

We made all scripts and tools available to let academia and industry evaluate and contribute to this work.

8.6 Conclusions and further research

In this study, we show the role of Linked Data and XBRL in bringing new opportunities for effective transparency in corporate reporting. Linked Data principles can encourage better corporate data publication and therefore data analysis, defining the interconnection across financial and non-financial data (such as sustainability data) and documents publicly available in open government data initiatives and voluntary reporting initiatives. XBRL enables a standard and accurate representation of corporate data with advanced validation rules. We present a solution to convert independent silos of XBRL data into interconnected pieces. Lessons learned during the process and benefits are presented. While our work demonstrates

the potential of this approach, it would benefit from extension in the following ways: (i) incorporate data sources beyond environmental, financial and DBpedia; (ii) incorporate non public-domain data, and address associated data protection issues necessary to do this; (iii) consider scalability and performance issues in the transformations necessary. In future work, we intend to evaluate and integrate sustainability reporting in XBRL, such as GRI data, and extend the ontology proposed using RDF Data Cube. We believe this study encourages scholars, regulators, data publishers and users to promote and use both XBRL and Linked Data, as each solution has a different role to play. Combined use of them enables non-financial factors such as environmental and social performance of companies to be integrated into reasoning, allowing improved transparency and accountability by a diverse group of stakeholders.

Chapter 9. A Solution To Aligning Corporate Reporting Frameworks: The Case of GRI and CDP

The previous chapters highlight the importance of increasing the connectivity between reporting frameworks to enhance the consideration of environmental information in decision-making processes. Chapter 8 presents a solution to connect datasets from different domains (financial and non-financial) and different formats. That technical solution is also valid for establishing relationships between similar data points from different datasets (e.g., Revenue in CNMV must be equal to the Revenue in DBpedia). However, the main difficulty is identifying these corresponding points, which is more an accounting problem than a technical issue. This chapter is trying to answer our fourth objective “How to support the consistency, analysis and data alignment between different reporting frameworks?”.

This chapter aims to present a methodology to facilitate the identification of common data points from different reporting frameworks using text mining techniques, as well as the market needs and opportunities. To demonstrate the usefulness of our proposal, we have tested our method using CDP and GRI datasets.

This work was accepted and presented at the 1st Data Amplified Conference - Academic track (Singapore, Malaysia). In addition, these ideas were used by the author to write several chapters of the book *Big Data and Business Information* (ISBN: 978-84-16286-32-4), led by the Bank of Spain and published by AECA.

9.1 Introduction

In response to the 2008 financial crisis, the European Union (EU) has sought to encourage firms to be more transparent in their activities. Transparency is considered an essential component to achieving a sustainable economy, enabling a clearer and open disclosure of information to tackle the economic, social and environmental challenges of the new millennium [112]. As a result, the European Union is enforcing better reporting practices, as set out in the recent European Directive 2014/95/EU for non-financial reporting. As of 2017, this applies to public interest organizations with more than 500 employees, which represents approximately 6000 large companies in Europe. This Directive defines a set of information to

be disclosed about social, environmental and financial risks matters, and suggests a set of voluntary reporting frameworks to enable compliance [113].

Currently financial and non-financial reporting regulations and voluntary standards are some of the mechanisms used to encourage the disclosure of financial, environmental and social impacts. The content of corporate reporting has evolved from primarily regulated reports on financial aspects, to a more multidimensional batch of compulsory and voluntary documents, containing both indicators and explanations of social, environmental and corporate governance behaviour. This information is mainly influenced by reporting initiatives based on guidelines for voluntary disclosures, such as the Global Reporting Initiative (GRI) and the CDP. The forthcoming EU regulation is going to put further pressure on businesses to disclosure risks, as well as, social and environmental impacts. It means that companies will need to move from voluntary to compulsory reporting regarding non-financial information, and they will need to care more about the quality, comparability and reliability of that information. Our central thesis is that better alignment between voluntary reporting frameworks can make a significant contribution to achieving part of the European objectives, and enhance the disclosure of social, environmental and financial risks. To that extent, we explore how GRI and CDP could align their areas of reporting, and what potential benefits are associated with such alignment.

GRI and CDP are recognised standard setters in Europe and worldwide, in sustainability information and environmental reporting respectively. In 2014, 1633 small and large EU organizations used the GRI framework to produce their sustainability reports [134], and 1365 EU companies disclosed climate data to CDP [135], illustrating that GRI and CDP have already achieved inroads with European companies. Of the 1633 GRI reporting firms in Europe, 670 companies use CDP guidelines to report environmental information. This equates to 41% of GRI reporting companies in the EU that disclose GHG (greenhouse gas) emissions and climate change strategies through CDP as well.

Since 2010, both initiatives joined efforts to align common areas of their reporting frameworks. Since then, every year CDP and GRI publish a document together called “Linking

GRI and CDP⁴⁸, where the cross-references between both frameworks are described in detail. That document specifies the correspondence between areas of reporting, but it still lacks perfect equivalence, which is a barrier to achieving part of the European objectives. Table 18 shows an example extracted from the Linking document, where the level of alignment between both frameworks related to Climate Change Risk is presented. The first column identifies the CDP questions that address that area, and the second column shows the correspondence with GRI concepts. The third column on the right makes clear that one concept in CDP is rarely an exact equivalent to one or more elements in GRI, and vice-versa; rather it is either broader or narrower in scope.

CC5. CLIMATE CHANGE RISK		
<p>CC5.1 Have you identified any inherent climate change risks that have the potential to generate a substantive change in your business operations, revenue or expenditure?</p> <p>[tick-box selection]</p>	<p>GENERAL STANDARD DISCLOSURES</p> <p>G4-2 <i>Extracts from G4-2-a:</i> a. Provide a description of key impacts, risks, and opportunities.</p> <p>[...]</p> <p>Section One should [...] include:</p> <ul style="list-style-type: none"> • A description of the significant economic, environmental and social impacts of the organization, and associated challenges and opportunities. This includes the effect on stakeholders' rights as defined by national laws and the expectations in internationally recognized standards and norms <p>[...]</p> <p>Section Two should include the following:</p> <ul style="list-style-type: none"> • A description of the most important risks and opportunities for the organization arising from sustainability trends 	<p>G4 reporters would report information requested by CC5.1 – CC5.1f under G4-2 and G4-EC2. However, G4-2 has a broader scope than the corresponding CDP questions, referring to economic, environmental and social issues more generally.</p>
<p>CC5.1a Please describe your inherent risks driven by changes in regulation</p> <p>[table question]</p>		
<p>CC5.1b Please describe your inherent risks that are driven by change in physical climate parameters</p> <p>[table question]</p>		
<p>CC5.1c Please describe your inherent risks that are driven by changes in other climate-related developments</p> <p>[table question]</p>		

Table 18. Convergence in the area of Climate Change risks between CDP and GRI frameworks

⁴⁸ Linking GRI and CDP (2016): How are GRI'S G4 guidelines and CDP'S 2016 climate change questions aligned? <https://goo.gl/JTYwxc>

We consider that aligning key facts from CDP and GRI reporting frameworks promote the following benefits:

- (1) more direct comparability of reports;
- (2) cross-validation between CDP and GRI facts;
- (3) reducing reporting burden: disclosing once and submitting twice to both CDP and GRI.

The purpose of the present study is to propose a methodology to determine the alignment between different reporting frameworks using text mining analysis. This involves identifying corresponding points between key terms contained in GRI and CDP reports. The results are of interest not only to scholars, but also for company filers and regulators in a context of growing integration between financial and non-financial reporting frameworks in Europe.

9.2 Problem situation

The problem addressed in this chapter belongs to the area of research called semantic integration, which is active in the use of databases, information-integration and ontologies. It is also gaining popularity in the Artificial Intelligence community as a way of facilitating knowledge sharing. Semantic integration is normally a process composed of three tasks [114]:

- 1) **Mapping discovery:** Given two domains of information, how do we find similarities or equivalences between them? This is a similar approach to LIME, the tool we used in Chapter 8 for discovering the links with other data sources in the web.
- 2) **Representation of mappings:** Once the mapping of similarities and equivalences are identified, how do we represent them to enable reasoning?
- 3) **Reasoning with mappings:** Once the mappings are represented, they are used for data transformation, query answering, web-service composition.

Regarding mapping discovery, the main goal is to map data structures identifying types of correspondence. Prior work on this area points to two different approaches to address this task:

- (1) **Extending ontologies** for general purposes with specific properties and concepts agreed upon by the developers. These types of ontologies are called Upper ontologies,

which are developed for that purpose; examples of these types of ontologies are the Suggested Upper Merged Ontology (SUMO) and Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE).

(2) **Using heuristics and machine learning methods** to find mappings using the characteristics of ontologies, such as definition of concepts, properties and structures.

In this study we focus on the mapping discovery task, identifying the level of equivalences between CDP and GRI frameworks, classifying the content of the unstructured GRI reports in CDP questions using Natural Language processing (a type of machine learning⁴⁹), and determining which concepts and properties represent similar notions to enable more direct comparability, cross-validation and less reporting burdens (which are our reasoning goals).

Our problem then is how to classify text under the given categories using automatic text categorisation. Text categorisation systems generally use a vector model representation of the documents. The vector that represents the document contains the document terms and also the weights assigned to each term. There have been many text classifiers proposed using machine learning techniques: neural networks, genetic algorithms and probabilistic models.

- **Probabilistic**, such as the Naïve Bayesian (NB). The basic idea in NB approaches is to use the joint probabilities of words and categories to estimate the probabilities of categories given a document. The main limitations of this type of classifier are the sensitivity to term reduction, and the difficulty of interpreting and maintaining the numeric models.
- The **Decision Tree** classifier is composed of nodes, branches and leafs. Nodes are labelled by terms, branches are labelled by the weight that the term has in the text document and leaves are labelled by categories [115]. It is built using a 'divide and conquer' strategy; in other words, this procedure checks whether all the examples have the same label and, if not, a new subtree is created from a term partitioning from the pooled classes of documents that have the same values.

⁴⁹ Machine learning grew from pattern recognition and the theory that computers can learn without being programmed to perform specific tasks.

- **Regression models** belong to the statistical-based class and are best known as linear least squares. This procedure models the data with two vectors: a vector for weighted terms and a vector of weighted categories. The classification consists of adjusting the category weights.
- **Neural Networks** consist of many individual processing units, the so-called neurons, connected by links, which can work in parallel. These links have weights associated with them and allow neurons to activate other neurons. A learning algorithm is used to find suitable weight values for the links [116]. A primary weakness of this method is the excessive training and computation time.
- **Support Vector Machine (SVM)** is defined over a vector space where the problem is to find a decision surface that best separates the data points in two classes [117]. In order to define the best separation, a margin between two classes is introduced. An interesting property of SVM is that the decision surface is determined only by the data points that have exactly the same distance from the decision plane. Those points are called the support vectors, which are the only useful elements in the training set; if all other points were removed, the algorithm would learn the same decision function. This property makes SVM theoretically unique and different from methods such as Neural Networks and Naïve Bayesian where all the data points in the training set are used to optimise the decision function.

Some studies [118][119] suggest that SVM significantly outperforms the other classifiers for the micro-level performance on pooled category assignments, which is why this method was selected to address the problem. However, it is out of the scope of this study to examine which classifier could be better and why. The main purpose is to propose a methodology to solve an alignment problem between reporting frameworks, and illustrate how Natural Language Processing (NLP) techniques can help to achieve this goal.

9.4 Method step by step

The first step is generating the classification model and carrying out training. To do so, we collected 23324 CDP responses⁵⁰ classified into six categories, which are the six questions related to the Business Strategy section of the Climate Change questionnaire (Figure 35). This data represents three years of historical data (2013-2015) from 2145 companies. This data is structured and was provided by CDP in Excel format.

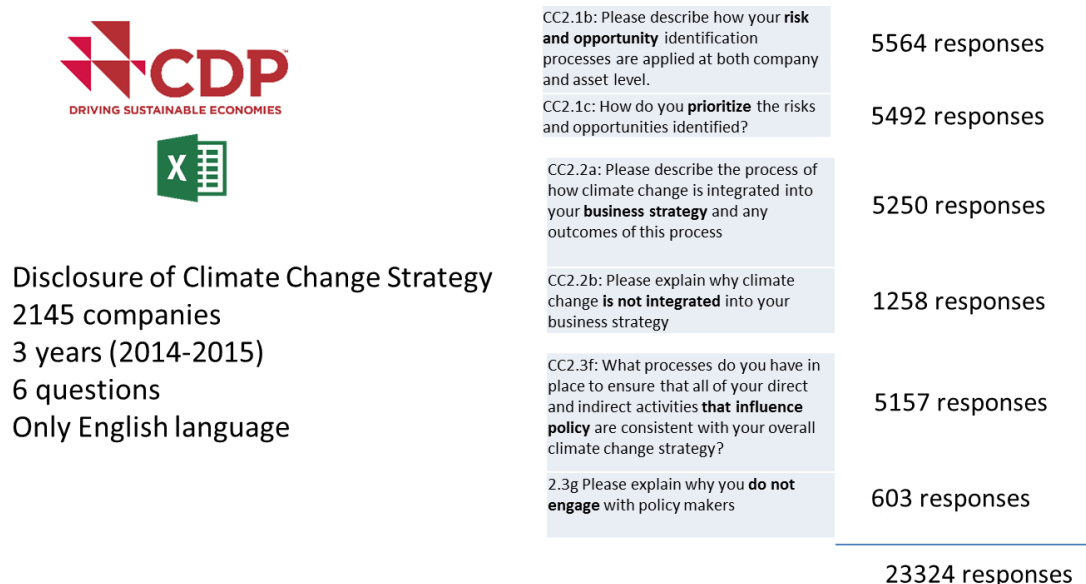


Figure 35. CDP responses classified by question number

It was essential to clean the text to make the text mining analysis more efficient [17]. This was done by removing punctuation, numbers, and whitespaces, and converting all text to lower case. It is also important to stem the words to retrieve their radicals and to remove the affixes to reduce noise from the texts. For that purpose, we used the suffix-stripping algorithm, which consists of inflected forms and root form relations. A smaller list of "rules" is stored to find the root form of each word; for example, if the word ends in 'ing', 'ed' or 'ly', the algorithm removes those parts and keeps the root. We applied the stemming process to all the CDP data normalised in plain text and applied the cleaning process described above.

CDP data was kindly provided by CDP for the purpose of this study.

We then proceeded to **generate the model** using Support Vector Machine (SVM) algorithm [18]. This algorithm plots each piece of CDP data in 2-dimensional space (Figure 36), where n is the number of features. In this case, n is the question number that each response belongs to. The value of each feature is the value of a coordinate. We then performed classification by finding the hyperplane. We used 80% of the CDP responses to train the model.

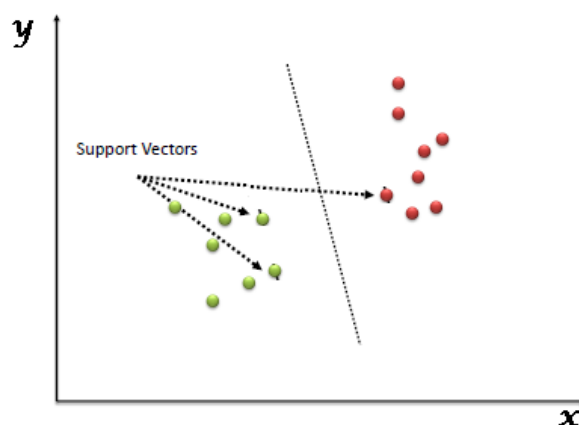


Figure 36. CDP data in 2-dimensional space

In order to evaluate whether the model generated is accurate, we followed [19] and used the remaining 20% of the training data for testing purposes:

- **Precision** is the number of true positives divided by the number of expected similarities.
- **Recall** is the number of similarities divided by the number of both similarities and differences.
- **Percentage** of overlap, obtained by means of the division of the size of the intersection of the two sets by the scale of the union of the sets and multiplying by 100.

The result in Table 19 shows that the model is suitable to classify most questions. These results were validated using cross-validation and verified by User Acceptance Testing (UAT) carried out the CDP Technical Director and myself.

	SVM_PRECISION	SVM_RECALL	SVM_FSCORE
1	0.88	0.85	0.86
2	0.83	0.83	0.83
3	0.92	0.84	0.88
4	0.58	0.39	0.47
5	0.73	0.91	0.81
6	0.73	0.44	0.55

Table 19. SVM results: precision, recall and score

Once the model was generated and evaluated, we proceeded to **classify the GRI data**. To do so, we used ten GRI reports that are publicly available on the GRI website⁵¹. These reports are unstructured data in PDF format with heading and sub-headings with no correlation with CDP reports, and each of them follows a different structure and design. They were normalised in plain text to facilitate the analysis using text mining techniques.

We proceeded to **clean and stem** the GRI data using the same techniques as in CDP data; each page/text from each report was then classified using the model. As a final step, we expect to classify each text/page into a CDP question. We cannot go to a level of classification such as “the GRI (5.1) data point is equal to CDP (4.2) data point” because there is no unique identifier per data point in the GRI reports.

This work was developed by using R software ⁵² and

- the Text Mining package ⁵³ for applying text mining functionalities required such as stemming and stopword removal.
- The e1071 package ⁵⁴ for using the SVM data classification method.

⁵¹ <http://database.globalreporting.org/>

⁵² R software: <https://www.r-project.org/>

⁵³ Text Mining package in R: <https://cran.r-project.org/web/packages/tm/tm.pdf>

⁵⁴ The e1071 package: <https://cran.r-project.org/web/packages/e1071/index.html>

9.5 Results

In total, 975 pages from ten GRI reports were classified in six different CDP questions. An example of how the results are obtained is presented in Table 20. If we interpret the first row, it means that the content (GRI_TEXT) of page 31 from the GRI report number 25 is classified as the CDP question number CC2.2a with a confidence level of 0.996 (out of 1).

DOC	PAGE	SVM_PROB	CDP_CATEGORY	GRI_TEXT
25	31	0.996380627	CC2.2a_BUSINESS_STRATEGY_CLIMATE_CHANGE	In addition to
25	26	0.991984381	CC2.2a_BUSINESS_STRATEGY_CLIMATE_CHANGE	
25	25	0.991773656	CC2.2a_BUSINESS_STRATEGY_CLIMATE_CHANGE	America to do
25	30	0.988395512	CC2.2a_BUSINESS_STRATEGY_CLIMATE_CHANGE	PILLAR 4
25	24	0.986643486	CC2.2a_BUSINESS_STRATEGY_CLIMATE_CHANGE	Air Canada
25	28	0.949205134	CC2.2a_BUSINESS_STRATEGY_CLIMATE_CHANGE	Green Aviation
25	29	0.930827209	CC2.2a_BUSINESS_STRATEGY_CLIMATE_CHANGE	yielded
25	15	0.907619585	CC2.3f_POLICY_MAKER_ENGAGEMENT	common
25	41	0.907305561	CC2.3f_POLICY_MAKER_ENGAGEMENT	
25	36	0.892748222	CC2.2a_BUSINESS_STRATEGY_CLIMATE_CHANGE	offset portal
25	19	0.876495421	CC2.3f_POLICY_MAKER_ENGAGEMENT	With the
25	34	0.864170972	CC2.3f_POLICY_MAKER_ENGAGEMENT	WORKING
25	35	0.840176947	CC2.3f_POLICY_MAKER_ENGAGEMENT	Waste
25	27	0.839161197	CC2.2a_BUSINESS_STRATEGY_CLIMATE_CHANGE	price of US\$106
25	4	0.829662692	CC2.3f_POLICY_MAKER_ENGAGEMENT	AT A GLANCE
25	10	0.816134856	CC2.1b_RISK_OPPORTUNITY_COMPANY_ASSET_LEVEL	GOVERNANCE
25	14	0.789427389	CC2.1c_PRIORITIZE_RISK_OPPORTUNITIES	To ensure that
25	33	0.728042953	CC2.2a_BUSINESS_STRATEGY_CLIMATE_CHANGE	model or age,
25	42	0.672492461	CC2.2a_BUSINESS_STRATEGY_CLIMATE_CHANGE	
25	17	0.658503415	CC2.1b_RISK_OPPORTUNITY_COMPANY_ASSET_LEVEL	observations or
25	43	0.604386422	CC2.3f_POLICY_MAKER_ENGAGEMENT	Canadian
25	21	0.602885243	CC2.3f_POLICY_MAKER_ENGAGEMENT	vacuum suction

Table 20. Results: GRI reports classified in CDP questions

Overall, as demonstrated in Figure 37, the highest confidence level belongs to questions CC2.2a and CC2.3f. The question CC2.1 likewise has a high level of trust even with fewer pages.

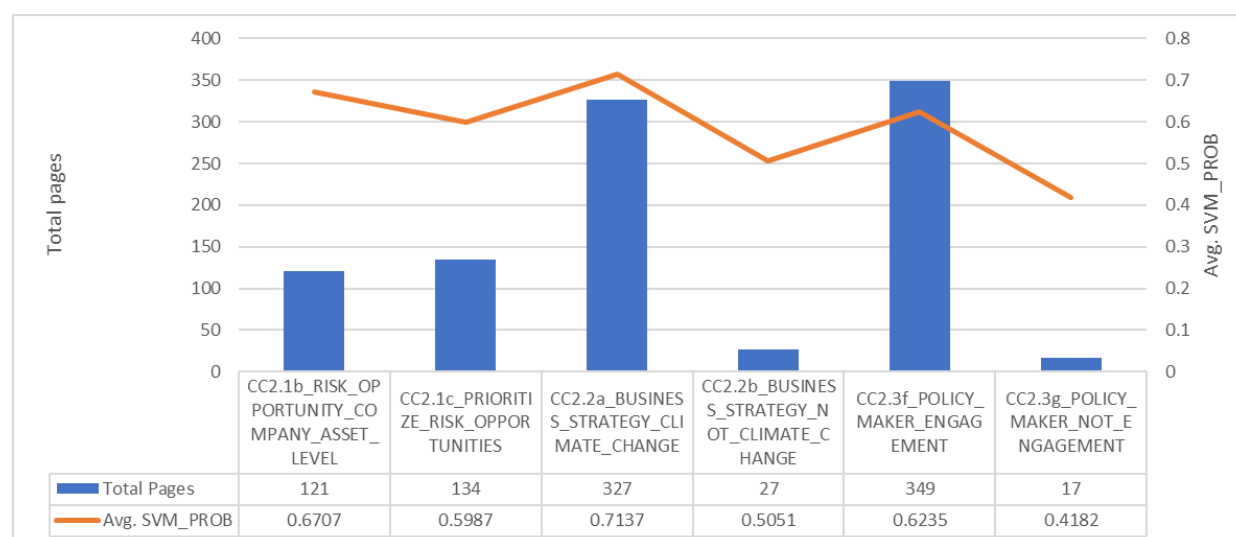


Figure 37. Total pages of GRI reports classified per CDP question number and confidence levels

Around 200 pages have a confidence level greater than 0.9, and around 400 with more than 0.7. Only 260 pages have less than 0.5. These results can be found in Figure 38.

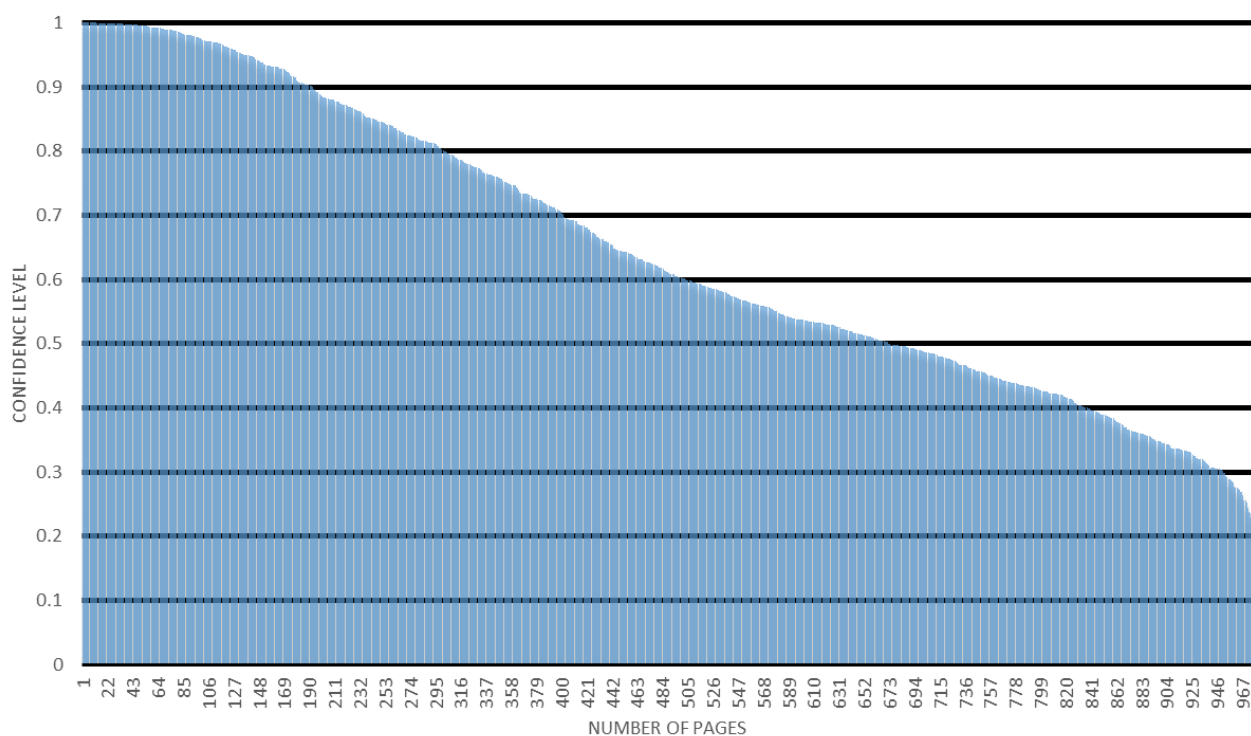


Figure 38. Number of GRI pages and associated confidence level

9.6 Discussion and conclusions

This chapter proposes a text mining methodology to determine alignments between GRI and CDP frameworks, and demonstrates that our method is able to discover relationships between different reporting frameworks, specifically:

- $CDP(\text{questions}) = GRI(\text{text per page})$ when confidence level greater than 0.9
- $CDP(\text{questions}) \sim GRI(\text{text per page})$ when confidence level greater than 0.7

We used the CDP data as a dictionary to build and test the SVM algorithm. This data was obtained in a structured format, and each piece of information was uniquely identified by question number. This is in opposition to GRI reports, which have an unstructured format and each key performance indicator is not uniquely identified. This represents a limitation and is the main reason why our method only can find relationships between CDP questions and GRI text pages, instead of CDP question and GRI indicator.

Given that one of the main needs in forthcoming regulations for corporate reporting is improving alignment between reporting frameworks, we believe that our solution is a necessary tool to facilitate the identification of equivalent information. Clear benefits are:

- **Analysis:** More direct comparability of GRI and CDP reports.
- **Quality:** Cross validations between GRI and CDP text information.
- **Reducing reporting burden:** Disclose once and submit twice to CDP and GRI.

Comparing GRI and CDP reporting practices, we would like to remark the following insights for further works:

- CDP reports are available centrally from CDP to those who are willing to pay; this does not occur with GRI, although some of the reports can be found on the GRI website. However, the common practice is that reports are published on companies' websites.
- Each data point in CDP reports is uniquely identified and it follows a key-value approach. This is the primary reason it was possible to build a dictionary. GRI reports follow a marketing approach, which means that each piece of information is not uniquely identified and they are presented without much precision in a qualitative manner. There is little evidence provided.
- CDP reports are structured information; different reports from different companies share the same container of information but with different content. Each GRI report is completely different; there is not a common container and reports are presented in a readable format (pdf) instead of a consumable format (e.g. CSV, XBRL, XML).

We believe this methodology can be useful:

1. To CDP and GRI: when they are evaluating and analysing the information that they share and how they share it. The use of methodologies like this during this process should encourage the improvement of the data quality and alignment between GRI and CDP framework.
2. To reporting companies that are disclosing to CDP and are using GRI frameworks to develop their annual reports, sustainability reports, etc.: this work should help to reduce the reporting burdens, allowing them to disclose once and submit twice.
3. To software providers that build reporting solutions: this methodology should help to incorporate assessment capabilities within their software.

4. To the users of this information, such as investors and governments: they can automatically compare certain data from CDP and GRI reports as well as have more confidence in the reliability of that information.

For further research, it would be interesting to explore the performance of other classifiers for this particular problem, as well as the different level of aggregations between reporting frameworks such as $CDP(x) = GRI(y) + GRI(z)$. For that purpose, the use of a clustering method instead of classification could be appropriate; in such a case it is necessary to group instances on the basis of common features.

Chapter 10. Conclusions and further work

Our research journey started with an investigation into the area of environmental sustainability reporting. The main issue is that corporate reporting practices are becoming increasingly complex, in part due to the necessity of extending the dialogue from shareholders and customers to other interest groups or wider stakeholders, such as employees, governments, analysts, scholars, and NGOs. As a result, a large amount of corporate information is being generated and published in an impractical way. For example, companies cannot compare their CO₂ intensity figure against their competitors, and investors cannot carry out a reliable risk analysis on a company's financial performance while also considering their social and environmental actions.

Yet companies defend themselves by saying that they are producing vast quantities of corporate information on both a voluntary and compulsory basis as a matter of transparency, meaning that they would like their actions to be analysed beyond financial issues. Governments are producing regulations to make the disclosure of non-financial information compulsory and investors are supporting reporting initiatives like CDP. Unfortunately, forthcoming regulations in this field, like the EU Directive detailed in Chapter 9, fail regarding practical details. The European Commission feels that they are proposing how information can be requested effectively to create a significant impact on decision-making processes for companies and their shareholders. They are still exploring why this extra financial information is relevant, instead of figuring out *how* they could require information in order to make a genuine impact on decisions. The reality on the ground is that companies are producing non-financial information on a voluntary basis that is of low quality and reliability; instead, it is used by the companies as a marketing and reputational tool, rather than to influence decision-making processes.

During the course of this study, we wanted to change the way companies work to improve the way they and their stakeholders make decisions. Decisions that look beyond financial matters are necessary if solutions are to be found for environmental challenges. To that purpose, businesses need to transform their management and operations, and information is critical to ensure a better integration of environmental issues within their decision-making process.

In sustainability reporting, our study focuses on the environmental aspects supported by Information Systems, specifically highlighting reporting standards, semantic technologies and Big Data, and exploring CDP as a case of study. Overall, we have responded to the following four research questions:

- 1) Market understanding on the use and disclosure of environmental data.
- 2) How to make environmental data more accessible for users?
- 3) How to enable the use of environmental data in combination to financial information from companies ?
- 4) How to support the consistency, analysis and data alignment between different reporting frameworks?

We also wanted to keep the balance between academic theories and pragmatism given that, as part of the EngD programme requirements and my personal interest, this research should reduce the gap between academia and industry, and must create impact in both worlds. This approach drove the direction of each stage of this study: from having a good understanding of the problem situation; asking the market and exploring relevant academic theories and prior works; envisioning new opportunities and ways to tackle pending challenges along with pilot solutions; and combining novel technologies approaches, such as Linked Data, XBRL and Big Data.

The results produced during this research take the form of several projects and collaborations, plus more than 20 publications and presentations in academic and industrial forums. They represent an incubator of ideas for further developments in academia and industry.

Three important aspects have influenced the direction of this project:

- 1) The opportunity of having CDP as the industrial sponsor and the guidance and support of my industrial supervisor, Pedro Faria (Technical Director at CDP). To me, CDP represented an opportunity to explore a real environmental reporting market with customers, competitors, missions and business model.

- 2) The opportunity of being part of an EngD programme and the guidance and support of my academic supervisor, Chris Preist (Reader in Sustainability and Computer Systems at the University of Bristol).
- 3) My 13 years of professional experience implementing technologies and XBRL in the financial arena (The Bank of Spain, the Spanish Security Exchange Commission, the Institute of Accounting and Accounts Auditing, General Intervention Board of the State, The Audit Chamber of Andalusia) and non-financial reporting domains (The Spanish Association of Accountants (AECA) and CDP).

The contributions of this study are summarised in the following sections.

10.1 The problem situation and vision

Throughout Chapters 4 and 5, we explored the vision of the ideal environmental reporting system agreed upon by all active stakeholders. These results were translated into market needs and demands, and academic target areas for further exploration.

Using stakeholders with different backgrounds, we identified what questions to ask, classified common needs, and identified which needs are particular to different stakeholders. During this evaluation, it was important to separate stakeholders into two groups: producers and consumers of information.

One of the most relevant discoveries in the course of this research is the drivers of environmental disclosure practices. This is the “why” behind non-financial disclosure, and they are classified as: regulatory compliance, stakeholder communication and reputation, competitive advantage and internal efficiency. These drivers generate two different demands:

- 1) More qualitative and explanatory information to better communicate their story and value creation in the short-, medium- and long-term to a diverse, multi-stakeholder group: customers, regulators, investors, employees and general society.
- 2) An evolution of non-financial information into structured, homogeneous and digital data to be comparable and easy to analyse.

From our point of view, it is fundamental to find the balance between story-telling and a value-creation approach so that this information is able to be analysed more easily.

When companies are in the practice of disclosing non-financial information, there is a problem regarding the different guidelines to follow. There is misalignment in the basic principles (i.e. thresholds of materiality), especially because of the variety of guidelines available to create such reports. For example, there are sustainability reports, Integrated reports, Corporate Social Responsibility Reports, Management reports and Director's reports. In the case of environmental information, CDP brings a certain level of uniformity and standardisation in the environmental domain. However, there is a demand for a unique standard to guide companies in preparing and disclosing financial and non-financial information in an integrated way, or at least reduce the lack of complementarity between the different reporting standards. Chapter 6 provides a qualitative study of the challenges around interoperability (technical, semantic, organisational, legal) to the exchange of environmental and sustainability data/information for better accountability (e.g. to policy and regulation) and decision making (e.g. within business and at policy level) when it comes to sustainability. We envision that supporting these claims using specific ICT solutions, such as XBRL and Linked Data, will allow new opportunities to support the areas of data standardisation, data connectivity, data accessibility and data integration.

An additional problem discovered is how reporting processes are carried out inside the organisation. For example, in the case of financial statements, it generally requires a coordinated effort between different departments, led by the financial department. The annual accounts must then be signed by the board of directors, and finally audited by an external auditor in order to ensure that the information is an accurate representation of the company and is consistent, reliable and free from material misstatement, whether caused by fraud or error. The auditors make sure that the financial statements present conformity with the generally accepted accounting principles. The audit report offers a reasonable level of assurance, which increases stakeholder confidence. In some cases, like in Spain and the United Kingdom, this information must be submitted in XBRL format to a government agency,

such as Companies House in the UK and The Business Registers in Spain. In both cases, the annual financial accounts are publicly available⁵⁵⁵⁶ by these authorities in XBRL format.

On the other hand, the non-financial reports like CSR/Sustainability/Integrated reports require coordinated work between different departments but, according to the result of our interviews, this does not always happen; normally the responsibility lies with the marketing or sustainability department. This also depends on the sector; for example, in oil and gas, companies normally have specific departments for the disclosure of environmental information, given that environmental decisions and actions are integrated into a company's core processes. These reports sometimes follow a review process instead of an audit process. The review process is usually carried out by an external auditor, which validates whether the information is consistent and only offers a limited level of assurance. However, unlike the auditing processes, the non-financial information lacks a generally accepted framework that makes impossible to ensure, with a reasonable level of assurance, that the information accurately represents the company. In Figure 39, the differences between the three levels of assurance—compilation, review and audit—are described in more detail.

⁵⁵ Companies House accounts data: <https://beta.companieshouse.gov.uk/> [Accessed August 2019]

⁵⁶ Business Register accounts data: <https://www.registradores.org/registroonline/home.seam> [Accessed August 2019]

	Compilation	Review	Audit
Assurance as to Generally Accepted Accounting Principles	None	Limited	Statements are fairly presented in accordance with GAAP
Entities Covered	Nonpublic	Nonpublic only	Public or nonpublic
Understanding of internal controls and assess control risk	Not required	Not required	Required
Inquiry procedures required	Not required	Inquiry and analytical procedures	Inquiry and analytical procedures
GAAP disclosures omitted	May be omitted provided accountant's report discloses omission	May be omitted provided accountant's report discloses omission	Inadequate disclosures requires qualified or adverse opinion

Figure 39. Levels of assurance in auditing. Source: <https://goo.gl/Ka94Tz> [Accessed August 2019]

In addition to the limited level of assurance, this information is normally not signed by the board, which explains why investors and other users of this information do not have the same level of trust as they do in financial disclosures. In addition, non-financial reports are usually published on corporate websites in PDF format and in an unstructured way. Unlike the financial statements, these reports are decentralised, unstructured, not audited and not signed by the board. In the case of CDP, the required environmental data is more structured, and CDP offers a place where this information is centralised. However, in general terms this means that the non-financial information is harder to analyse than the financial reports and less reliable with regards to making decisions.

Figure 40 illustrates the different levels of assurance through an example of an annual report published by Triodos Bank. An annual report is a comprehensive report on a company's activities and developments. It is published yearly and sent to the company's shareholders and other stakeholders. This report includes part of the financial reports and non-financial reports, such as the income statement, the balance sheet and CO2 emissions, as well as

information on the events that have influenced the company's accounting through the years. Each part of this report offers a different level of assurance.

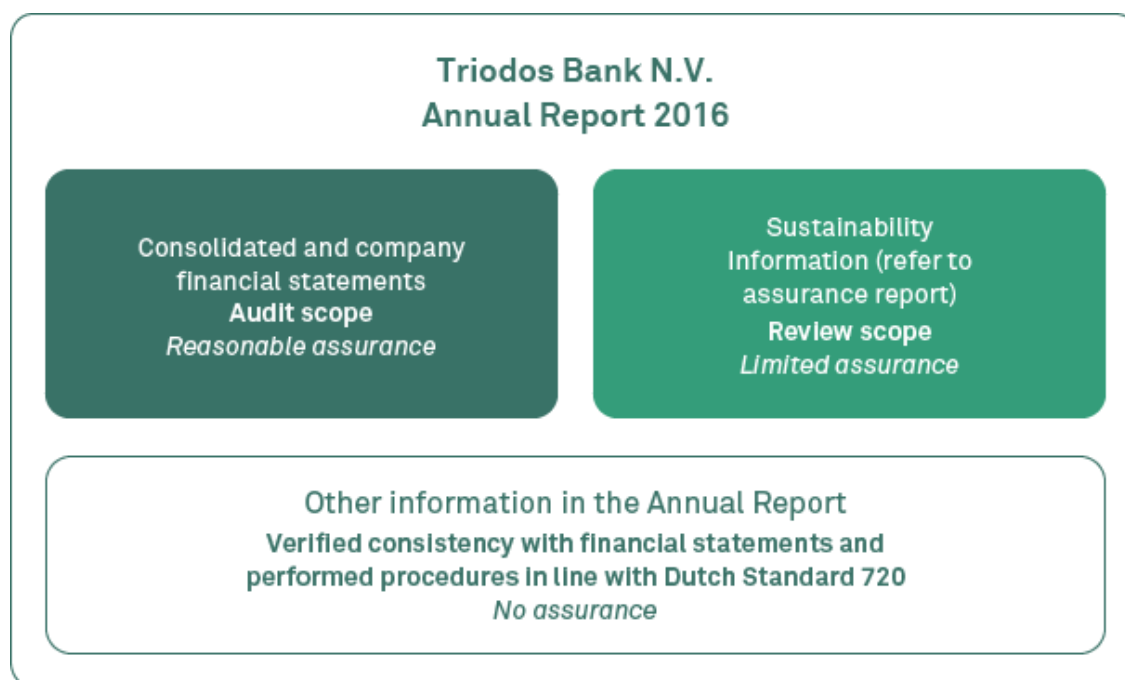


Figure 40. Annual report and the different level of assurance

In addition to how reporting processes are established inside companies and the different levels of assurance, there are several accounting problems in the non-financial domain.

- Lack of precision, accuracy and reliability in non-financial metrics. This produces gaps in analysis and validation.
- Lack of auditing and assurance processes/methodologies.
- Lack of analysis by combining information from different domains, e.g., what is the financial impact of an increase in carbon emissions?

Non-financial reporting initiatives like CDP are sharing stakeholders and reporting problems, and, to some extent, these problem areas are already solved in the financial reporting domain. We recommend that non-financial reporting initiatives such as CDP learn from the financial reporting practices and solutions implemented in related academic work. During our research, we took the approach that (1) evolving the XBRL standard (mainly in use in regulatory financial reporting) to include certain properties that are needed to represent and

validate an environmental information domain, and (2) implementing XBRL for environmental information using CDP data.

10.2 Solutions

This study used innovative solutions to address several technical problems (Figure 41):

- **Lack of data standardisation and data quality:** In Chapter 7, we envisioned the opportunities that XBRL could offer in climate-related reporting challenges. We demonstrated it is possible to: (1) develop the CDP data model in XBRL format; (2) evolve the XBRL standard through participation in the community as a member of the Best Practice Board at XBRL International; and (3) build the XBRL adoption strategy in CDP: <https://www.cdp.net/en/research/xbrl>.
- **Lack of accessibility and interoperability:** We developed a solution for publishing CDP data connected to existing financial datasets issued by supervisory authorities and other Linked Open Datasets found on the Web, and we brought more contextual information and levels of validation to these corporate datasets. This solution encourages CDP to not only publish new environmental data on the Web, but to link with existing reports. These ideas were first envisioned in Chapter 6 and implemented using the potential of XBRL and Linked Data in Chapter 8.
- **Lack of alignments between different reporting frameworks:** As previously discussed, greater complementarity is needed between the different reporting standards. In Chapter 9, we presented Big Data methodology using text mining techniques to evaluate the level of alignments or equivalences between CDP and GRI data points, considering the real practices in corporate reporting analysing the company reports. Chapter 9 demonstrates how these techniques can be used for this particular reporting problem, as well as the continuing challenge of addressing how the unstructured information is provided in a decentralised way without unique identification or homogeneity.
- **Lack of proper tools for visualisation and analysis:** We did not cover solutions to this subject, but it could be a future area of exploration. We recommend approaching this area with a multidisciplinary group formed from experts in environmental science and accounting, and researchers on computational climate models who are familiar with a data science approach.

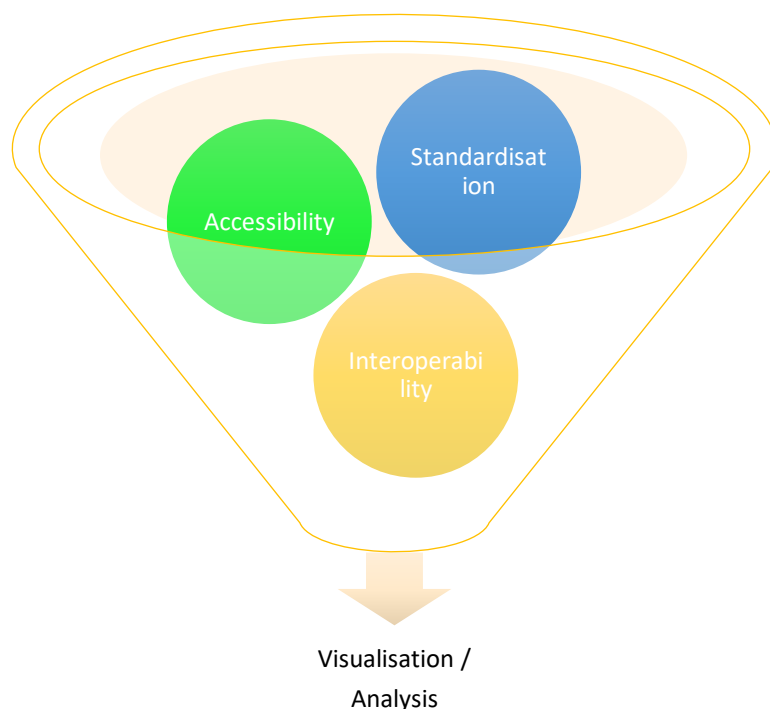


Figure 41. Technical problems in environmental reporting

10.3 Impact and contributions of this research

Our research has contributed to academic knowledge through solution-driven papers, book chapters, journal papers and presentations at academic conferences. We have identified and articulated the continuing challenges and opportunities in environmental reporting for a multi-stakeholder audience. We believe that our contributions provide a fundamental platform to integrate both technologies and communities more effectively, and therefore enhance corporate transparency.

Regarding industry, I have gained the position to be the Chair of the Best Practices Board ⁵⁷at XBRL International, Expert Member of the International Financial Reporting Standards Board, Expert member of the ISO TC322 for Sustainable Finance and TC68 Financial Services and Chair the European XBRL Academic conference, hosted by the European Central Bank. This

⁵⁷ Best Practices Board at XBRL Int: <https://www.xbrl.org/the-consortium/get-involved/best-practices-board/>
[Accessed August 2019]

academic conference represents a space for sharing academic ideas and industrial needs, and it is helping to create new industrial networks and academic opportunities.

Regarding our industrial impact, it takes the form of:

- New collaborations with organisations addressing similar reporting challenges, such as AECA, CDP and the IIRC, but whose work is at the national and global level.
- Practical influence in global standards organisations such as XBRL International by evolving the standard to support reporting needs from the environmental information domain.
- Active participation in EU meetings regarding reporting policies, as well as a participative role in reporting working groups led by the ICAEW, FRC, AECA, XBRL International, XBRL Europe and XBRL Spain.

Overall, this research has represented a personal and professional challenge with a fascinating learning process and collaborative network. The results provide an exploration of the problem space and innovative solutions that can be further developed by industry and academia. Through these results, I wanted to inspire others and bring a certain level of evolution and “out of the box thinking” into traditional corporate reporting environments. I hope this EngD research will aid the work being done at CDP, other sustainability initiatives and those following with related academic work.

10.4 Further work

Given the exploratory and solution-driven nature of this work, it can serve as a resource for further research into environmental information systems and multidisciplinary work.

We believe that the challenges identified at the production, consumption and distribution of environmental sustainability information should be addressed through a joint effort between business strategy, accounting and technology as presented in Figure 42.

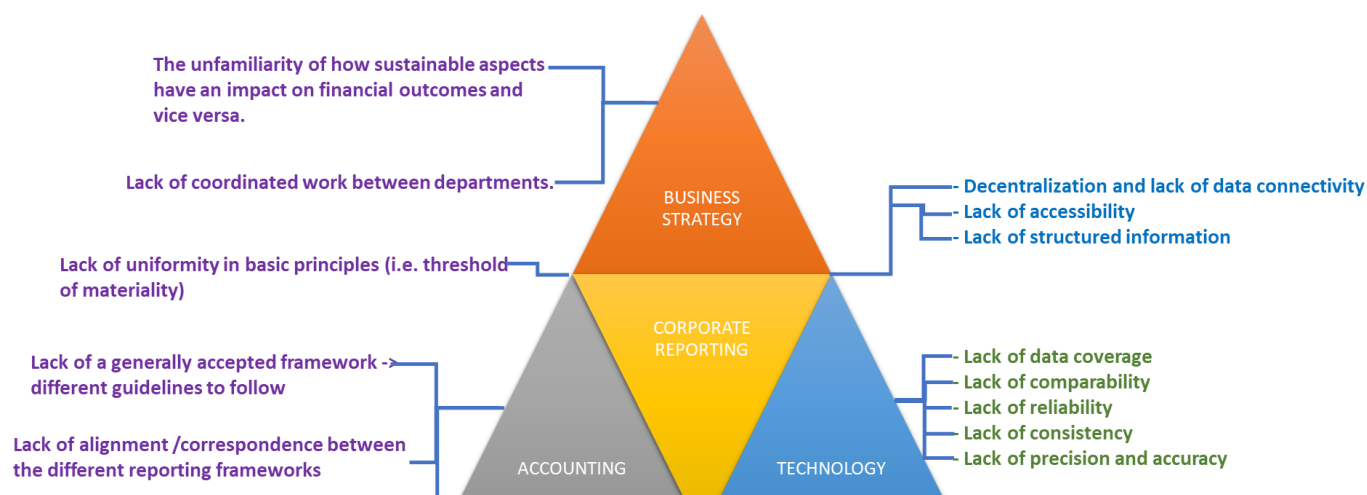


Figure 42. Disciplines that can solve corporate reporting problems

Our work should inspire the further exploration and development in this field:

- From an accounting perspective, it would be valuable to identify how to create a Generally Accepted Framework for the disclosure of financial and non-financial information that could promote a higher level of assurance, in line with efforts by Lizcano *et al.* [120].
- From a business strategy perspective, it is important to identify what business goals are necessary to achieve in the context of the real impact of ESG information [121]. Apart from identifying the market-relevant information, such an investigation will provide a better understanding of the importance of the types of disclosure (e.g. environmental, social or governance) and the best way to present the information (e.g. whether qualitative or quantitative). It is important to explore the reality of the reporting, studying and comparability of company and sector-specific disclosure best practices against the requirements of global and national regulatory bodies (e.g. GRI, IRC, FASB, IASB, UK FRC, and various reporting frameworks) in a manner that is consistent with fundamental principles such as materiality [122].

- It would be valuable to explore which short-term and long-term effects could be derived from the “connected-models” application of certain transparency policies [123].
- Finally, what kind of decisions could be driven by interoperability, given the corporate reporting ecosystem? We believe that our study can facilitate the identification of design patterns to build better IS solutions for sustainability, in line with previous work such as Knowles *et al.* [124][125]. For example, assessing the format, content and quality of all forms of corporate disclosure on ESG matters to let this diverse information be examined via topic modelling, text mining and other Artificial Intelligence techniques.

11.References

- [1] United Nations (2019). Environmental Sustainability definition available at: <https://bit.ly/2KmqfiN> [Accessed August 2019]
- [2] Unerman, J. (2000) Methodological issues-reflections on quantification in corporate social reporting content analysis. *Accounting, Auditing & Accountability Journal*, 13(5), 667-681.
- [3] Ihlen, J. and Roper, J. (2014). Corporate reports on sustainability and sustainable development: 'We have arrived'. *Sustainable development*, 22(1), 42-51.
- [4] IEEE (2017). Available at <http://bit.ly/1RyeAfE> [Accessed August, 2019]
- [5] Azzone, G. Bertele, U. (1994) Exploiting green strategies for competitive advantage. *Long Range Planning* 17(7), 69–81.
- [6] Peloza, J. (2009). The Challenge of Measuring Financial Impacts from Investments in Corporate Social Performance. *Journal of Management* 35(6), 1518-1541.
- [7] Salzmann, O., Ionescu-Somers, A., and Steger, U. (2005). The Business Case for Corporate Sustainability: Literature Review and Options. *European Management Journal*, 23(1):27–36.
- [8] Myers, M., and Newman M. (2007). The qualitative interview in IS research: Examining the craft. *Information and organization* 17.1 (2007): 2-26.
- [9] Hannabuss, S. (1996). Research interviews. *New Library World*, 97(5), 22-30.
- [10] Walsham, G. (1995). Interpretive case studies in IS research: nature and method. *European Journal of information systems*, 4(2), 74-81.

- [11] Schultze, Ulrike, and Michel Avital. (2011). Designing interviews to generate rich data for information systems research. *Information and Organization* 21.1 : 1-16.
- [12] O'Leary, C., Rao, S. and Perry, C. (2004). Improving customer relationship management through database/Internet marketing: A theory-building action research project. *European journal of marketing*, 38(3/4), 338-354.
- [13] Jasperson, J., Carte, T., Saunders, C., Butler, B., Croes, H. and Zheng, W. (2002). Review: power and information technology research: a metatriangulation review. *MIS quarterly*, 26(4), 397-459.
- [14] Patton, M. Q. (1990). Qualitative evaluation and research methods. *SAGE Publications, inc.*
- [15] Chawla, N. V., Hall, L. O., Bowyer, K. W., and Kegelmeyer, W. P. (2004). Learning ensembles from bites: A scalable and accurate approach. *The Journal of Machine Learning Research*, 5, 421-451.
- [16] Caruana, R. and Niculescu-Mizil, A. (2006). An empirical comparison of supervised learning algorithms. *In Proceedings of the 23rd international conference on Machine learning* (pp. 161-168). ACM.
- [17] Stavrianou, A., Andritsos, P. and Nicoloyannis, N. (2007). Overview and semantic issues of text mining. *ACM Sigmod Record*, 36(3), 23-34.
- [18] Vapnik, V. *Statistical Learning Theory*. Wiley, New York, NY, (1998).
- [19] Good, B., and Tennis, J. T. (2009). Term based comparison metrics for controlled and uncontrolled indexing languages. *Information research*, 14(1), 8.

- [20] Rezaee, Z. (2017). Business sustainability: Performance, compliance, accountability and integrated reporting. *Routledge*.
- [21] EU Commission (2014). Directive 2013/34/EU as regards disclosure of non-financial and diversity information by certain large undertakings and groups. Available at <http://goo.gl/NWZGD2> [Accessed August 2019]
- [22] Maffini C., Marques J., Kruglianskas I., Aparecida L. and Schoproni R. (2015). Management for sustainability: An analysis of the key practices according to the business size. *Ecological Indicators* 52: 116-127.
- [23] Ramos T, Martins I., Martinho A., Douglas, Painho M., Caeiro S. (2014). An open participatory conceptual framework to support state of the environment and sustainability reports *Journal Cleaner Production*, 64 (1) , pp. 158-172
- [24] Gupta, S., and Goldar, B., (2005). Do stock markets penalize environment-unfriendly behaviour? Evidence from India. *Ecological Economics*, 52(1), 81-95. <http://dx.doi.org/10.1016/j.ecolecon.2004.06.011>
- [25] Wahba, H., (2008). Does the market value corporate environmental responsibility? An empirical examination. *Corporate Social Responsibility and Environmental Management*, 15(2), 89-99. <http://dx.doi.org/10.1002/csr.153>
- [26] STOXX (2011). Transparency and flexibility stand out in the flood of 'sustainable' products. Available at: <http://bit.ly/PGCqGo> [Accessed August 2019].
- [27] Bansal, Pratima. (2005). Evolving sustainably: A longitudinal study of corporate sustainable development." *Strategic management journal* 26.3: 197-218.
- [28] Lazonick, W. and O'sullivan, M. (2000). Maximizing shareholder value: a new ideology for corporate governance. *Economy and society*, 29(1), 13-35.

- [29] Kumar N. and Singh J. P. (2013). Effect of board size and promoter ownership on firm value: some empirical findings from India. *Corporate governance*, 13(1).
- [30] EBR (2018). EBR framework available at: <https://goo.gl/T82Jb5> [Accessed August 2019].
- [31] GRI (2017). Available at: <http://database.globalreporting.org/> [Accessed August, 2019].
- [32] Eleftheriadis, I. and Anagnostopoulou, E. G. (2015). Identifying barriers in the diffusion of renewable energy sources. *Energy Policy*, 80, 153-164.
- [33] AECA (2019) Integrated Scoreboard guidance of reporting available at <https://bit.ly/2MOVNwF> [Accessed August 2019]
- [34] Jansson, A., Nilsson, F., and Rapp, B., (2000). Environmentally driven mode of business development: A management control perspective. *Scandinavian Journal of Management*, 16(3), 305-333. [http://dx.doi.org/10.1016/S0956-5221\(99\)00022-6](http://dx.doi.org/10.1016/S0956-5221(99)00022-6)
- [35] Gartner. (2012). Big data. Glossary 2012. Available at: <https://goo.gl/trHLzU> [Accessible August, 2019]
- [36] Chaudhuri, P. P. (1998). Business practices: the quiet revolution. *Span (July/August)*, 3-6.
- [37] Brobst, S. and Rarey J. (2003). Five Stages of Data Warehouse Decision Support Evolution. *DSSResources. COM* 1, no. 06
- [38] European Commission (2010). Communication from the commission to the European Parliament, the council, the European Economic and social committee and the committee of The Regions: Towards interoperability for European public services. European Commission, vol. COM, 12/16/2010, pp. 1–55.

- [39] European Commission (2015). Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, A Digital Single Market Strategy for Europe. European Commission, COM,2015/192 final.
- [40] Lee, D., Hosanagar, K. and Nair, H. (2014). The effect of social media marketing content on consumer engagement: Evidence from Facebook. *Stanford Graduate School of Business*.
- [41] Molloy, J. C. (2011). The Open Knowledge Foundation: Open Data Means Better Science. *Plos Biology*, 9, 1-4.
- [42] Mitchell, I., and Wilson, M. (2012). Linked Data: Connecting and exploiting big data. *White paper. Fujitsu UK*, 302.
- [43] Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R. and Hellmann, S. (2009). DBpedia-A crystallization point for the Web of Data. *Web Semantics: science, services and agents on the world wide web*, 7(3), 154-165.
- [44] Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., and Ives, Z. (2007). Dbpedia: A nucleus for a web of open data. In *The semantic web* (pp. 722-735). Springer, Berlin, Heidelberg
- [45] UK Open dataData Gov, (2018). Website available at: <https://data.gov.uk/> [Accessed August 2019]
- [46] Watson, R. T., Boudreau, M. C. and Chen, A. J. (2010). Information systems and environmentally sustainable development: energy informatics and new directions for the IS community. *MIS quarterly*, 23-38.
- [47] Anitha, J., and Aruna M. (2014). Adoption of Human Resource Information System in Organisations. *Journal of Contemporary Research in Management* 9.4: 63.

- [48] Herzig, Christian, and Stefan S. (2006). Corporate sustainability reporting. An overview. *Sustainability accounting and reporting*. Springer Netherlands, 301-324.
- [49] Silvius, A. and Ron S. (2012) Sustainability in the business case. In *Proceedings of the 26th IPMA World Congress, Crete, Greece*, vol. 2931.
- [50] Machado, C., de Lima, E. P., da Costa, S. E. G., Angelis, J. J., and Rocha, L. M. (2016). Studying sustainability process implementation through operations management lens. *KTH, Industriell Management*.
- [51] Lux, W. (2014). Integrated Performance Management strategy, risk and sustainability Management. *Review of business and economics studies*.
- [52] Leviston, Z., Leitch A., Greenhill M., Leonard M., and Walker I. (2011). Australians' views of climate change. *Canberra: CSIRO* (2011).
- [53] Comyns, B. and Figge, F. (2015). Greenhouse gas reporting quality in the oil and gas industry: A longitudinal study using the typology of "search", "experience" and "credence" information. *Accounting, Auditing & Accountability Journal*, 28(3), 403-433.
- [54] CDP. (2017). Disclosers companies at CDP available at: <https://www.cdp.net/en/companies-discloser> [Accessed August 2019]
- [55] Wegener, M., Elayan, F. a., Felton, S., and Li, J. (2013). Factors Influencing Corporate Environmental Disclosures. *Accounting Perspectives*, 12(1), 53–73. doi:10.1111/1911-3838.12007
- [56] Beder, S. (2006) Suiting themselves. How corporations drive the global agenda. *Earthscan*: London.

- [57] Tang, Q. and Luo, L. (2014). Carbon Management Systems and Carbon Mitigation. *Australian Accounting Review*, 24(1), 84–98. doi:10.1111/auar.12010
- [58] Andrew, J. and Cortese, C. (2011). Accounting for climate change and the self-regulation of carbon disclosures. *Accounting Forum*, 35(3), 130–138. doi:10.1016/j.accfor.2011.06.006.
- [59] Wegener, M., Elayan, F. a., Felton, S., and Li, J. (2013). Factors Influencing Corporate Environmental Disclosures. *Accounting Perspectives*, 12(1), 53–73. doi:10.1111/1911-3838.12007
- [60] Song, W., and Szewczyk, S. (2003). Does coordinated institutional investor activism reverse the fortunes of underperforming firms?. *Journal of financial and quantitative analysis*, 38(2), 317-336.
- [61] Rankin, M., Windsor, C., and Wahyuni, D. (2011). An investigation of voluntary corporate greenhouse gas emissions reporting in a market governance system: Australian evidence. *Accounting, Auditing and Accountability Journal*, 24(8), 1037–1070. doi:10.1108/09513571111184751
- [62] Stanny, E. and Ely, K. (2008). Corporate environmental disclosures about the effects of climate change. *Corporate Social Responsibility and Environmental Management*, 15(6), 338-348.
- [63] Sullivan, R., and Gouldson, A. (2012). Does voluntary carbon reporting meet investors' needs? *Journal of Cleaner Production*, 36(January), 60–67. doi:10.1016/j.jclepro.2012.02.020
- [64] Wegener, M., Elayan, F. a., Felton, S., and Li, J. (2013). Factors Influencing Corporate Environmental Disclosures. *Accounting Perspectives*, 12(1), 53–73. doi:10.1111/1911-3838.12007

- [65] Lydenberg S. (2010). How to Read a Corporate Social Responsibility Report: A User's Guide. *Boston College Centre for Corporate Citizenship*.
- [66] Johnston, D., Sefcik, S. and Soderstrom, N. (2008). The value relevance of greenhouse gas emissions allowances: An exploratory study in the related United States SO₂ market. *European Accounting Review*, 17(4), 747-764.
- [67] Gallego-Álvarez, I., Rodríguez-Domínguez, L. and García-Sánchez, I. (2011). Study of some explanatory factors in the opportunities arising from climate change. *Journal of Cleaner Production*, 19(9-10), 912–926. doi:10.1016/j.jclepro.2011.02.012
- [68] Kolk, A., Levy, D. and Pinkse, J. (2008). Corporate Responses in an Emerging Climate Regime: The Institutionalization and Commensuration of Carbon Disclosure. *European Accounting Review*, 17(4), 719–745. doi:10.1080/09638180802489121
- [69] Strong, D. M., Lee, Y. W., & Wang, R. Y. (1997). Data quality in context. *Communications of the ACM*, 40(5), 103-110.
- [70] Eccles, R. G., Ioannou, I. and Serafeim, G. (2014). The impact of corporate sustainability on organizational processes and performance. *Management Science*, 60(11), 2835-2857.
- Watson
- [71] Lam, Aris YC, Mei Mei Lau, and Ronnie Cheung.(2016).Modelling the Relationship among Green Perceived Value, Green Trust, Satisfaction, and Repurchase Intention of Green Products." *Contemporary Management Research* 12.1: 47.
- [72]Melville, N. P. (2010). Information systems innovation for environmental sustainability. *MIS quarterly*, 34(1), 1-21.
- [73] Scott, W. R. (1987). The adolescence of institutional theory. *Administrative science quarterly*, 493-511.

- [74] Adela, Chen JW, Marie-Claude Boudreau, and Richard T. Watson. (2008). Information systems and ecological sustainability. *Journal of Systems and Information Technology* 10.3: 186-201.
- [75] Diez, E. and McIntosh, B. S. (2011). Organisational drivers for, constraints on and impacts of decision and information support tool use in desertification policy and management. *Environmental Modelling and Software*, 26(3), 317-327. doi: 10.1016/j.envsoft.2010.04.003
- [76] Ahmed, M. D. and Sundaram, D. (2012). Sustainability modelling and reporting: From roadmap to implementation. *Decision Support Systems*, 53(3), 611–624. doi:10.1016/j.dss.2012.02.004.
- [77] Bonson-Ponte, E., Torres, L., Royo, S. and Flores-Munoz, F. (2012). Local e-government 2.0: Social media and corporate transparency in municipalities." *Government information quarterly* 29, no. 2 (2012): 123-132.
- [78] Searcy, C.;Elkhawas, D.(2012). Corporate sustainability ratings: an investigation into how corporations use the Dow Jones Sustainability Index. *Journal of Cleaner Production* 35, 79-92.
- [79] Cohen, M. and Santhakumar, V. (2007). Information disclosure as environmental regulation: a theoretical analysis. *Environmental and Resource Economics*, 37(3), 599-620.
- [80] Eccles, R. G., and Krzus, M. P. (2010). *One report: Integrated reporting for a sustainable strategy*. John Wiley & Sons.
- [81] Vasarhelyi, M. A., Alles, M., Kuenkaikaew, S. and Littlely, J. (2012). The acceptance and adoption of continuous auditing by internal auditors: A micro analysis. *International Journal of Accounting Information Systems*, 13(3), 267-281.

- [82] Arndt, H., Isenmann, R., Brosowski, J., Thiessen, I., and Gómez M. (2006). Sustainability reporting using the extensible business reporting language (XBRL). *Managing environmental knowledge*, 75-82.
- [83] Hřebíček, J., Štencl, M., Trenz, O., and Soukopová, J. (2011). Corporate performance evaluation and reporting. In *Proc. International Conference on Environment, Economics, Energy, Devices, Systems, Communications, Computers, Pure and Applied Mathematics* (pp. 338-343).
- [84] Valentinetti, D. and Rea, M. A. (2011). Adopting XBRL in Italy: Early evidence of fit between Italian GAAP Taxonomy and current reporting practices of non-listed companies.
- [85] Bonson-Ponte, E., Cortijo, V. and Escobar, T. (2009). Towards the global adoption of XBRL using International Financial Reporting Standards (IFRS). *International Journal of Accounting Information Systems*, 10(1), 46-60.
- [86] Isenmann, R. and Gomez, M. (2009). Advanced corporate sustainability reporting–XBRL taxonomy for sustainability reports based on the G3-guidelines of the Global Reporting Initiative. In *Hřebíček, J. Towards eEnvironment. European Conference of the Czech Presidency of the Council of the EU*(pp. 25-27).
- [87] Troshani, I. and Rao, S. (2007). Drivers and inhibitors to XBRL adoption: A qualitative approach to build a theory in under-researched areas. *International Journal of E-Business Research (IJEER)*, 3(4), 98-111.
- [88] Janvrin, D. and Won G. (2012). XBRL implementation: A field investigation to identify research opportunities. *Journal of Information Systems* 26.1 : 169-197.
- [89] Vandenbussche, P. Y., Atemezing, G. A., Poveda-Villalón, M. and Vatan, B. (2017). Linked Open Vocabularies (LOV): a gateway to reusable semantic vocabularies on the Web. *Semantic Web*, 8(3), 437-452.

- [90] Yoose, B. and Perkins, J. (2013). The linked open data landscape in libraries and beyond. *Journal of Library Metadata*, 13(2-3), 197-211.
- [91] R  ther, M., Fock, J., H  bener, J., Bandholtz, T., and Schulte-Coerne, T. (2010). Linked Environment Data. In *EnviroInfo*(pp. 470-479).
- [92] Cayzer, S., and Preist, C. (2010). The sustainability hub: an information management tool for analysis and decision making. *ACM Sigmetrics Performance Evaluation Review*, 37(4), 84-88.
- [93] Garcia, A. and Gil, R. (2009). Publishing XBRL as Linked Open Data. *World Wide Web Workshop: Linked Data on the Web (LDOW2009)*, vol. 538
- [94] Kampgen, B., Weller, T., Oriain, S., Weber C. and Harth, A. (2014). Accepting the XBRL Challenge with Linked Data for Financial Data Integration. *European Semantic Web Conference*, pp. 595–610. Springer International Publishing
- [95] Madlberger, L., Tho  ni, A., Wetz, P., Schatten, A., Tjoa, A. (2013). Ontology-based Data Integration for Corporate Sustainability Information Systems. *Proceedings of International Conference on Information Integration and Web-based Applications*, ACM
- [96] Goto, M., Hu, B., Naseer, A. and Vandenbussche, P. (2013). Linked data for financial reporting. *Proceedings of the 4th International Conference on Consuming Linked Data*, vol. 1034, pp. 123–135, CEUR-WS
- [97] Elliot, S. (2011). Transdisciplinary perspectives on environmental sustainability: a resource base and framework for IT-enabled business transformation. *MIS Quarterly*, 35(1), 197-236.
- [98] Srour, M., Baird, K. and Schoch, H. (2016). The Role of Strategic Flexibility in The Associations between Management Control System Characteristics and Strategic Change. *Contemporary Management Research*, 12(3).

- [99] COP21. (2015) United Nations Conference of the Parties, 21st session; Framework Convention on Climate Change FCCC Adoption of the Paris Agreement, 2015, Paris: UN
- [100] Newell, S. and Marabelli, M. (2014). The crowd and sensors era: opportunities and challenges for individuals, organizations, society, and researchers. *AIS Electronic Library*.
- [101] Beyer and Laney D. (2012). The Importance of “Big Data”: A Definition, *Gartner report*. pp. 1–9.
- [102] Mezgár and U. Rauschecker. (2014). The challenge of networked enterprises for cloud computing interoperability. *Computers in Industry*, 65(4), 657-674.
- [103] Soltysik-Piorunkiewicz and J. Banasikowska. (2013). “Interoperability and Standardization of e-Government Ubiquitous Systems in the EU Member States”. European Conference on e-Government, p. 481
- [104] Mora-Gonzalbez, J., Mora-Rodriguez, M. (2012). XBRL and Integrated Reporting: The Spanish Accounting Association Taxonomy Approach. *The International Journal of Digital Accounting Research*, vol. 12, pp.59–91
- [105] Bonson-Ponte, E., Escobar-Rodriguez, T and Flores-Munoz, F. (2007). The role of metadata language implementation in the European banking supervision network. *International Journal of Networking and Virtual Organisations*, 4(3), 245-256.
- [106] Stefan, D. and Letier E. (2014) .Supporting Sustainability Decisions in Large Organisations. *In Proceedings of the 2014 conference ICT for Sustainability* (Vol. 2, pp. 53-69). Atlantis Press.
- [107] Lorenz M., and Aebischer B. (2015). ICT for sustainability: An emerging research field. *ICT Innovations for Sustainability*. Springer International Publishing, 3-36.

- [108] Grabot, B. and Schlegel T. (2014). ICT for sustainability in industry. *Computers in Industry* 65.3 (2014): 383-385.
- [109] TFCD (2016). Charter Report available at <https://goo.gl/12UK6h> [Accessed August 2019]
- [110] Hoffman C. and Mora-Rodriguez M. (2013). Digitizing Financial Reports – Issues and Insights: A Viewpoint. *The International Journal of Digital Accounting Research*, vol. 13: 73-98. http://dx.doi.org/10.4192/1577-8517-v13_3
- [111] Garcia, R. and Gil, R. (2010). Triplicating and Linking XBRL Financial Data. *Proceedings of the 6th International Conference on Semantic Systems, ACM*
- [112] Gibbs, L. M. (2006). Valuing water: variability and the Lake Eyre Basin, central Australia. *Australian Geographer*, 37(1), 73-85.
- [113] EU Commission (2014). Directive 2013/34/EU as regards disclosure of non-financial and diversity information by certain large undertakings and groups. Available at <http://goo.gl/NWZGD2> [Accessed August 2019]
- [114] Noy, N. F. (2004). Semantic integration: a survey of ontology-based approaches. *ACM Sigmod Record*, 33(4), 65-70.
- [115] Pawar, P. Y., and Gawande, S. H. (2012). A comparative study on different types of approaches to text categorization. *International Journal of Machine Learning and Computing*, 2(4), 423.
- [116] Jeschke, G. and Lalmas, M. (2002). Hierarchical text categorisation based on neural networks and Dempster-Shafer theory of evidence. In *EUROFUSE Workshop on Information Systems*.

- [117] Vapnik, V.N. (1982). Estimation of Dependences Based on Empirical Data, Addendum 1, New York: SpringerVerlag.
- [118] Shen, L., Jiaming L., Shengyin W., and Shiqi Y. (2014). HEp-2 image classification using intensity order pooling based features and bag of words. *Pattern Recognition* 47, no. 7 : 2419-2427.
- [119] Dalal, R., Ismail, S., Rajesh, P., Divya, R. and Vivek K. (2015). A Lexicon Pooled Machine Learning Classifier for Opinion Mining from Course Feedbacks. In *Advances in Intelligent Informatics*, pp. 419-428. Springer, Cham.
- [120] Lizcano, J.L., Flores Muñoz, F., and Rejón López, M. (2011). *El informe integrado: primeras aproximaciones de aplicación. El caso español*. Available at http://www.aeca.es/comisiones/rsc/estudio_empirico.pdf [Accessed August 2019]
- [121] Marquis C., Beunza D., Ferraro F., Thomason B. (2011). Driving Sustainability at Bloomberg L.P. Harvard Business School Case, 411–025
- [122] ICAEW (1999), *Implementing Turnbull: A Boardroom Briefing*, Centre for Business Performance, London.
- [123] Karp, J. A., Banducci, S. A. and Bowler, S. (2003). To know it is to love it? Satisfaction with democracy in the European Union. *Comparative Political Studies*, 36(3): 271–292.
- [124] Knowles, B., Blair, L., Walker, S., Coulton, P., Thomas, L. and Mullagh, L. (2014). Patterns of persuasion for sustainability. In *Proceedings of the 2014 conference on Designing interactive systems* (pp. 1035-1044). ACM.
- [125] Knowles, B., Clear, A. K., Mann, S., Blevis, E., and Håkansson, M. (2016). Design Patterns, Principles, and Strategies for Sustainable HCI. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 3581-3588). ACM.

- [126] Pawlowski, S. D., and Robey, D. (2004). Bridging user organizations: Knowledge brokering and the work of information technology professionals. *MIS Quarterly*, 645-672
- [127] Butler, T. (2011). Compliance with institutional imperatives on environmental sustainability: Building theory on the role of Green IS. *The Journal of Strategic Information Systems*, 20(1), 6-26.
- [128] Brown, D., Dillard, J. and Marshall, R. (2005). Strategically informed, environmentally conscious information requirements for accounting information systems. *Journal of Information Systems*, 19(2), 79-103.
- [129] Ide and J. Pustejovsky (2010). What does interoperability mean, anyway? Toward an operational definition of interoperability for language technology. Proceedings of the Second International Conference on Global Interoperability for Language Resources. Hong Kong, China.
- [130] Panetto and J. Cecil. (2013). "Information systems for enterprise integration, interoperability and networking: theory and applications". *Enterprise Information Systems*, 7(1), 1-6.
- [131] Nativi, P., Mazzetti, A. and Geller, G. (2013). Environmental model access and interoperability: The GEO Model Web initiative. *Environmental Modelling and Software*, 39, 214-228.
- [132] SEC (2018). Security and Exchange Commission and XBRL website available at: <http://xbrl.sec.gov> [Accessed August 2019]
- [133] European Commission (2016). Available at <https://goo.gl/7AqsZP> [Accessed August, 2019]

11. References

[134] GRI, (2016), GRI framework available at: <https://goo.gl/PcWY8h> [Accessed August, 2019]

[135] CDP. (2016). CDP Website available at: www.cdp.net [Accessed August 2019]

Annex A - Mandatory and voluntary disclosure initiatives

Reporting initiative	Type	Scope	Disclosers	Type of related information	Mandatory/Voluntary	Channel	Format	Publicly available	Frameworks referenced	Disclosure location
European Union (EU): EU Directive 2014/95 regarding disclosure of non-financial and diversity +A2:K9+A2:K7information (2014)	Government reporting requirements and guidelines	Europe	European public interest organisations with more than 500 employees (around 6000 large companies in the EU)	Business model Policies and due diligence Outcomes (how financial and non- financial outcomes relates) Opportunities and risks that affect the business Sectorial and geographical scope Thematic aspects (often inter-connected): Environmental matters Social and employee matters Respect for human rights Anti-corruption and bribery matters Supply chain aspects Conflict minerals	Mandatory	Management report or separate non-financial statement referenced in the management report	Undefined. It depends on local transpositions	Depends on local transpositions	CDP (formerly the Carbon Disclosure Project); Climate Disclosure Standards Board; OECD Due Diligence Guidance for Responsible Supply Chains from Conflict-Affected and High-Risk areas, and the supplements to it; Eco-Management and Audit Scheme (EMAS) and the related Sectoral Reference Documents; European Federation of Financial Analysts Societies' KPIs for Environmental, Social, Governance (ESG), a Guideline for the Integration of ESG into Financial Analysis and Corporate Valuation; Global Reporting Initiative; Guidance for Responsible Agricultural Supply Chains of FAO-OECD; Guidance on the Strategic Report of the UK Financial Reporting Council; Guidelines for Multinational Enterprises of the Organisation for Economic Cooperation and Development; Guiding Principles Reporting Framework on Business and Human Rights; ISO 26000 of the International Organisation for Standardisation; International Integrated Reporting Framework; Model Guidance on reporting ESG information to investors of the UN Sustainable Stock Exchanges Initiative; Natural Capital Protocol; Product and Organisation Environmental Footprint Guides; Sustainability Accounting Standards Board; Sustainability Code of the German Council for Sustainable Development; United Nations (UN) Global Compact; UN Sustainable Development Goals, Resolution of 25 September 2015 transforming our world: the 2030 Agenda for Sustainable Development; UN Guiding Principles on Business and Human Rights implementing the UN 'Protect, Respect and Remedy' Framework.	This Directive depends on local transposition and each member of state is the responsible of adapting the management report to comply with these new requirements for 2018. For example, in Spain the Security and Exchange Commission (CNMV) will be responsible to receive this information.
Japan - Ministry of Environment	Government reporting pilot requirements	Japan	Any Japanese company	There are no concrete KPIs, just open disclosure of: Environmental Social Governance	Voluntary	ESG report Corporate Environmental Report	Submission after registration and access to a web portal. Report is submitted or linked. Many formats are permitted: PDF, XBRL, SNS etc.	Yes		https://www.env-report.env.go.jp/
Spain - Management report	Government reporting requirements	Spain	Public interest organisations with more than 500 employees	It is now changing and must align with the type of information required by the EU Directive 2014/95	Mandatory	Management report is part of the annual report	PDF	Yes	CII-FESG Framework (AECA) GRI UNCTAD	CNMV data portal: https://www.cnmv.es/Portal/Consultas/IFA/ListaIdFA.aspx?id=0&nif=A-08209769
TCFD (Task Force on Climate-related Financial Disclosure)	Guidelines	Global (G20)	G20: Listed and non-listed companies of any size. Sector oriented	Climate-Related Risks and Potential Financial Impacts	Voluntary	Undefined	Undefined	Undefined	Australia: National Greenhouse and Energy Reporting Act (2007) European Union (EU): EU Directive 2014/95 regarding disclosure of non-financial and diversity information (2014) France: Article 173, Energy Transition Law (2015) India: National Voluntary Guidelines on Social, Environmental, and Economic Responsibilities of Business (2011) United Kingdom: Companies Act 2006 (Strategic Report and Directors' Report) Regulations 2013 United States: NAICS, 2010 Insurer Climate Risk Disclosure Survey United States: SEC Guidance Regarding Disclosure Related to Climate Change Australia: Australia Securities Exchange Listing Requirement 4.10.3; Corporate Governance Principles and Recommendations (2014) Brazil: Stock Exchange (BM&FBovespa) Recommendation of report or explain (2012) China: Shenzhen Stock Exchange Social Responsibility Instructions to Listed Companies (2006) Singapore: Singapore Exchange Listing Rules 711A & 711B and Sustainability Reporting Guide (2016) ("Guide") South Africa: Johannesburg Stock Exchange Listing Requirement Paragraph 8.63; King Code of Governance Principles (2009) World, regional, and country-specific indices: S&P Dow Jones Indices, Sustainability Index, Sample Questionnaires Asset Owners Disclosure Project 2017 Global Climate Risk Survey CDP Annual Questionnaire (2016) GRESB Infrastructure Asset Assessment & Real Estate Assessment GRI Sustainability Reporting Standards (2016) IIGCC Oil & Gas (2010) Automotive (2009) Electric Utilities (2008) IIRC International Integrated Reporting Framework (2013) IPIECA Oil and gas industry guidance on voluntary sustainability reporting PRI Reporting Framework (2016) SASB Conceptual Framework (2013) and SASB Standards (Various)	
CDP (Carbon Disclosure Project)	Reporting System Guidelines	Global	Listed and non-listed companies of any size. Sector specific oriented. Cities	Environmental: Climate Change, Water and Forest	Voluntary	CDP Disclosure platform	Web platform- manual submission	The full dataset is available after payment and only part of the data is free.	GHG Protocol Science-based Targets Call to Action: RE100 CDSB Climate Policy Carbon Pricing (UN Global Compact)	CDP systems (the full dataset is available after payment); CDP open data portal (subset of the data that is free): https://data.cdp.net

Annex A - Mandatory and voluntary disclosure initiatives

GRI (Global Reporting Initiatives)	Guidelines	Global	Listed and non-listed companies of any size. Sector specific oriented.	Economic, Social, Environmental and Governance.	Voluntary	Annual reports ESG reports Sustainability reports Integrated Reports	PDF and XBRL	Yes	<p>International Labour Organization (ILO) Convention 87, 'Freedom of Association and Protection of the Right to Organise Convention', 1948.</p> <p>International Labour Organization (ILO) Convention 98, 'Right to Organise and Collective Bargaining Convention', 1949.</p> <p>International Labour Organization (ILO) Convention 135, 'Workers' Representatives Convention', 1971.</p> <p>International Labour Organization (ILO) Convention 154, 'Collective Bargaining Convention', 1981.</p> <p>International Labour Organization (ILO) Declaration, 'Declaration on Fundamental Principles and Rights at Work', 1998.</p> <p>International Labour Organization (ILO), Key Indicators of the Labour Market (KILM)</p> <p>International Labour Organization (ILO), LABORSTA Internet</p> <p>International Labour Organization (ILO) Recommendation 91, 'Collective Agreements Recommendation', 1951.</p> <p>International Labour Organization (ILO) Recommendation 163, 'Collective Bargaining Recommendation', 1981.</p> <p>International Labour Organization (ILO), 'Resolution concerning the International Classification of Status in Employment (ICSE)', 1993.</p> <p>Organisation for Economic Co-operation and Development (OECD) Principles, 'Principles of Corporate Governance', 2004.</p> <p>United Nations (UN), Composition of macro geographical (continental) regions, geographical sub-regions, and selected economic and other groupings</p> <p>United Nations (UN) Declaration, 'The Rio Declaration on Environment and Development', 1992.</p> <p>United Nations (UN), 'Guiding Principles on Business and Human Rights, Implementing the United Nations "Protect, Respect and Remedy" Framework', 2011.</p> <p>United Nations (UN), 'Protect, Respect and Remedy: a Framework for Business and Human Rights', 2008.</p> <p>United Nations (UN), Report of the Special Representative of the Secretary-General on the Issue of Human Rights and Transnational Corporations and Other Business Enterprises, John Ruggie, 2011</p>	The centralization of these reports is not compulsory. Companies normally publish their reports on their websites. GRI makes some of them publicly available on their website as a collection of PDFs: http://database.globalreporting.org/
IIRC (International Integrated Reporting Council)	Guidelines	Global	Listed and non-listed companies of any size	<p>Financial Capital</p> <p>Manufactured Capital</p> <p>Intellectual Capital</p> <p>Social and relationship capital</p> <p>Human capital</p> <p>Natural capital</p>	Voluntary	Integrated Reports Annual reports	PDF	Yes		The centralization of these reports is not compulsory. Companies normally publish their reports on their websites. IIRC makes a repository of some of them publicly available as a collection of PDFs: http://examples.integratedreporting.org/home
EMAS (Eco-Management and Audit Scheme)	Reporting System Guidelines	Europe	European public and private organisations	<p>Key performance indicators in six key environmental areas:</p> <p>Energy efficiency</p> <p>Material efficiency</p> <p>Water</p> <p>Waste</p> <p>Biodiversity</p> <p>Emissions</p>	Voluntary	EMAS report	PDF	Yes		EMAS repository: http://ec.europa.eu/environment/emas/register/